

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:08:29 ; Search time 46 Seconds

(without alignments)
1535.507 Million cell updates/sec

Title: US-10-339-656-2

Sequence: 1 MCFLEASHSVEDTPSHY.....QRTTDEALHPFDLKKK 445

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq.19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2410	100.0	445	23	AAE20170	Human kinase prote
2	2294.5	95.2	481	22	AAE20170	Human kinase prote
3	2177.5	90.4	479	19	AAE20170	Human kinase prote
4	2127.5	88.3	499	22	AAE20170	Human kinase prote
5	2117.5	87.9	451	20	AAE20170	Human kinase prote
6	1887	78.3	484	21	AAE20170	Human kinase prote
7	1808.5	75.0	482	19	AAE20170	Human kinase prote
8	1535	63.7	283	21	AAE20170	Human kinase prote
9	1526	63.3	350	21	AAE20170	Human kinase prote

10	1517	62.9	283	21	AAE20170	Human kinase prote
11	1400.5	58.1	352	23	AAE20170	Human kinase prote
12	1331.5	55.2	497	19	AAE20170	Human kinase prote
13	1314.5	54.5	490	22	AAE20170	Human kinase prote
14	1310.5	54.4	431	22	AAE20170	Human kinase prote
15	1223.5	50.8	485	19	AAE20170	Human kinase prote
16	1222	50.7	511	22	AAE20170	Human kinase prote
17	1222	50.7	511	22	AAE20170	Human kinase prote
18	1051	43.6	265	24	AAE20170	Human kinase prote
19	848.5	35.2	425	21	AAE20170	Human kinase prote
20	848.5	35.2	427	21	AAE20170	Human kinase prote
21	848.5	35.2	435	21	AAE20170	Human kinase prote
22	811.5	33.7	398	21	AAE20170	Human kinase prote
23	811.5	33.7	464	21	AAE20170	Human kinase prote
24	811.5	33.7	478	21	AAE20170	Human kinase prote
25	748	31.0	335	21	AAE20170	Human kinase prote
26	748	31.0	345	21	AAE20170	Human kinase prote
27	748	31.0	348	21	AAE20170	Human kinase prote
28	550.5	22.8	527	23	AAE20170	Human kinase prote
29	550.5	22.8	551	22	AAE20170	Human kinase prote
30	549.5	22.8	526	23	AAE20170	Human kinase prote
31	549.5	22.8	549	21	AAE20170	Human kinase prote
32	546.5	22.7	528	19	AAE20170	Human kinase prote
33	546.5	22.7	528	22	AAE20170	Human kinase prote
34	544.5	22.6	568	19	AAE20170	Human kinase prote
35	544.5	22.6	568	21	AAE20170	Human kinase prote
36	544.5	22.6	568	22	AAE20170	Human kinase prote
37	544.5	22.6	588	19	AAE20170	Human kinase prote
38	544.5	22.6	588	21	AAE20170	Human kinase prote
39	544.5	22.6	588	21	AAE20170	Human kinase prote
40	544.5	22.6	588	22	AAE20170	Human kinase prote
41	536.5	22.3	553	22	AAE20170	Human kinase prote
42	532.5	22.1	722	22	AAE20170	Human kinase prote
43	517	21.5	620	21	AAE20170	Human kinase prote
44	505	21.0	539	22	AAE20170	Human kinase prote
45	501.5	20.8	527	22	AAE20170	Human kinase prote

ALIGNMENTS

RESULT 1	AAE20170	standard; Protein; 445 AA.
ID	AAE20170	
AC	AAE20170	
XX	18-JUN-2002 (first entry)	
XX	Human kinase protein.	
DT	Human kinase protein.	
XX	Human kinase protein.	
DE	Human kinase protein.	
XX	Human kinase protein.	
KW	Human; kinase protein; enzyme; cytosolic; osteopathic; gene expression;	
KW	colon-moderately differentiated adenocarcinoma; chromosome mapping; gene;	
KW	drug screening; therapeutic; gene therapy; tissue typing; chromosome 5;	
KW	bone osteosarcoma.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Modified-site	10..13 "Casein kinase II phosphorylation site"
FT	Modified-site	24..31 "Tyrosine kinase phosphorylation site"
FT	Modified-site	25..28 "Tyrosine kinase phosphorylation site"
FT	Modified-site	29..36 "Casein kinase II phosphorylation site"
FT	Modified-site	55..61 "Tyrosine kinase phosphorylation site"
FT	Modified-site	69..71 "Tyrosine kinase phosphorylation site"
FT	Modified-site	72..74 "Protein kinase C phosphorylation site"
FT	Modified-site	72..74 "Protein kinase C phosphorylation site"

FT Modified-site 73..76 /note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 76..78 /note= "Protein kinase C phosphorylation site"
FT Modified-site 94..96 /note= "Protein kinase C phosphorylation site"
FT Modified-site 97..100 /note= "Protein kinase C phosphorylation site"
FT Modified-site 102..105 /note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 128..131 /note= "Casein kinase II phosphorylation site"
FT Modified-site 129..153 /note= "Casein kinase II phosphorylation site"
FT Binding-site 147..152 /note= "Protein kinase ATP-binding region signature"
FT Modified-site 176..179 /note= "N-myristoylation site"
FT Modified-site 209..212 /note= "Asn is N-glycosylated"
FT Modified-site 246..258 /note= "Casein kinase II phosphorylation site"
FT Active-site 247..250 /note= "Serine/threonine protein kinase active-site signature"
FT Modified-site 277..279 /note= "Casein kinase II phosphorylation site"
FT Modified-site 292..295 /note= "Protein kinase C phosphorylation site"
FT Modified-site 303..305 /note= "Casein kinase II phosphorylation site"
FT Modified-site 324..344 /note= "Protein kinase C phosphorylation site"
FT Region 368..370 /note= "Helix 1"
FT Modified-site 425..427 /note= "Protein kinase C phosphorylation site"
FT Modified-site 429..432 /note= "Protein kinase C phosphorylation site"
FT Modified-site /note= "Casein kinase II phosphorylation site"
XX MO200216567-A2.
XX
XX
PD 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US26389.
XX
XX 24-AUG-2000; 2000US-227470P.
XX PR 19-MAR-2001; 2001US-0810671.
XX
XX (APPL-) APPLERA CORP.
XX
XX Yan C, Ye J, Ketchum KA, Di Francesco V, Beasley EW,
PI WPI; 2002-269354/31.
XX DR N-PSDB; AAD32038.
XX
XX New human kinase proteins and nucleic acids, useful in drug screening
PT assays, identifying modulators of kinase activity or treating disorders
PT characterized by absence or unwanted expression of the protein
XX
XX Claim 1; Fig 2; 81pp; English.
XX
XX The invention relates to isolated human kinase proteins and nucleic
CC acids. The nucleic acid and peptide sequences can be used as models for
CC the development of human therapeutic targets, aid in the identification
CC of therapeutic proteins and serve as targets for the development of
CC human therapeutic agents that modulate kinase activity in cells and
CC tissues that express the kinase. The nucleic acids are useful as probes
CC or primers, in constructing recombinant vectors, for expressing
CC antigenic portions of the proteins, chromosome mapping, drug screening,
CC testing an individual for a genotype, and for gene therapy in patients

CC containing cells that are aberrant in kinase gene expression. The
CC proteins may be used in drug screening assays, in the identification of
CC compounds that modulate, stimulate or inhibit kinase activity, in
CC pharmacogenomic analysis, in treating disorders characterized by an
CC absence or unwanted expression of the protein (bone osteosarcoma, or
CC colon/moderately differentiated adenocarcinoma), and in generating
CC antibodies specific for the peptides. Such antibodies can be used to
CC detect the protein in situ, in vitro, or in cell lysate or supernatant,
CC to isolate and purify the proteins from host cells, pharmacogenomic
CC analysis, tissue typing, and in inhibiting protein function. The present
CC sequence is human kinase protein. Human kinase protein gene is
XX located on chromosome 5.

SQ Sequence 445 AA;

Query Match 100.0%; Score 2410; DB 23; Length 445;
Best Local Similarity 100.0%; Pred. No. 1,8e-221;
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCIPLEASHSVEDTTPSHYLEARSLNERDYDRRYVDEYRNDYCEGVYPRHYRDIESG 60
1 MCIPLEASHSVEDTTPSHYLEARSLNERDYDRRYVDEYRNDYCEGVYPRHYRDIESG 60
Db 1 MCIPLEASHSVEDTTPSHYLEARSLNERDYDRRYVDEYRNDYCEGVYPRHYRDIESG 60
QY 61 YRHCSSKSVSRSSPKRRNRHSCSHOSRSKSHRRKRSRS1EDDEBGLICQSGDVLK 120
61 YRHCSSKSVSRSSPKRRNRHSCSHOSRSKSHRRKRSRS1EDDEBGLICQSGDVLK 120
Db 61 YRHCSSKSVSRSSPKRRNRHSCSHOSRSKSHRRKRSRS1EDDEBGLICQSGDVLK 120
QY 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVKNGVYRRAASEIQLVHLNSTDP 180
121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVKNGVYRRAASEIQLVHLNSTDP 180
Db 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVKNGVYRRAASEIQLVHLNSTDP 180
QY 181 NSVFCVOMLEWMDHHGHCIVFELGLSTYDIKENSFLPFQIDHROMAYICQSGINF 240
181 NSVFCVOMLEWMDHHGHCIVFELGLSTYDIKENSFLPFQIDHROMAYICQSGINF 240
Db 181 NSVFCVOMLEWMDHHGHCIVFELGLSTYDIKENSFLPFQIDHROMAYICQSGINF 240
QY 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300
241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300
Db 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300
QY 301 LVSTRHYRAPEVILAGWSQPCDWSIGCILLEYLGFVFOYHDSKHLAMMERILGPI 360
301 LVSTRHYRAPEVILAGWSQPCDWSIGCILLEYLGFVFOYHDSKHLAMMERILGPI 360
Db 301 LVSTRHYRAPEVILAGWSQPCDWSIGCILLEYLGFVFOYHDSKHLAMMERILGPI 360
QY 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420
361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420
Db 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420
QY 421 EYDPTQRTILDALQHPFDLLKKK 445
421 EYDPTQRTILDALQHPFDLLKKK 445
Db 421 EYDPTQRTILDALQHPFDLLKKK 445

RESULT 2
AAB69705 standard; Protein; 481 AA.
XX ID AAB69705 standard; Protein; 481 AA.
XX
XX AAB69705;
XX
XX 14-AUG-2001 (first entry)
XX
XX Human cell cycle regulating protein 53.
XX
XX Human; cell cycle regulating protein 53; cancer; blood disease; HIV;
KW immunological disease; inflammation.
XX
XX Homo sapiens.
XX
XX WO200130833-A1.
XX
XX 03-MAY-2001.
XX
XX 16-OCT-2000; 2000WO-CN00328.
XX

PR 22-OCT-1999; 99CN-0119816.
 XX
 PA (SHAN-) SHANGHAI BIO ROAD GENE DEV LTD.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WPI: 2001-300480/31.
 DR N-PSDB; AAF99402.
 XX
 PT New cell cycle-regulating protein 53 and its polynucleotide, applicable
 PT in diagnosis and treatment of malignant tumour, haemopathy, human
 PT immunodeficiency virus infection, immunological diseases and various
 PT inflammation
 XX
 PS Claim 1: Page 23-24; 30pp; Chinese.
 XX
 CC The present invention provides the protein and coding sequences for human
 CC cell cycle regulating protein 53. The sequences can be used in the
 CC diagnosis and treatment of malignant tumours, haemopathy, human
 CC immunodeficiency virus (HIV) infection, immunological diseases and
 CC various types of inflammation. The present sequence is the cell cycle
 CC regulating protein 53.
 CC
 XX
 SQ Sequence 481 AA;
 Query Match 95.2%; Score 2294.5; DB 22; Length 481;
 Best Local Similarity 95.5%; Pred. No. 2.3e-210;
 Matches 429; Conservative 2; Mismatches 7; Indels 11; Gaps 2;
 QY 8 SH-SVEEDTH-----PSHYLEARSINERDYRDRRYVDEYRNDYCEGYVPRHYRD 56
 DB 33 SHSSTQENRHKCPHHQFKESDCHYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYRNG 92
 QY 57 IESGRIHCSKSSVSRSSSPKRNKRHCSSHQSSKSHRRKRSISIEDDEGHILICQSG 116
 DB 93 IESGYRIHCSKSSVSRSSSPKRNKRHCSSHQSSKSHRRKRSISIEDDEGHILICQSG 152
 QY 117 DVLARAYEIVDTLGGAGFGKVECTIDHGMDCGHVAVKIVKNGYRFAARSTIOVLEHNL 176
 DB 153 DVLARAYEIVDTLGGAGFGKVECTIDHGMDCGHVAVKIVKNGYRFAARSTIOVLEHNL 212
 QY 177 STDPSVFRVCVOMLEMFHDHGHVCIFFELLSGVDFIKENSFLLPFQIDHICOMAYOICQ 236
 DB 213 STDPSVFRVCVOMLEMFHDHGHVCIFFELLSGVDFIKENSFLLPFQIDHICOMAYOICQ 272
 QY 227 SINFLHNKLTHTTDLKPNILFVKSDYVVKYKSKMRKDRRTLKNTDIKAVDFGSAATYDDE 296
 DB 273 SINFLHNKLTHTTDLKPNILFVKSDYVVKYKSKMRKDRRTLKNTDIKAVDFGSAATYDDE 332
 QY 297 HHSITVSTRHVARPEVILALGMSQPCDWSIGCILLIYVIGFTVFQTHDSKHLAMMERI 356
 DB 333 HHSITVSTRHVARPEVILALGMSQPCDWSIGCILLIYVIGFTVFQTHDSKHLAMMERI 392
 QY 357 LGPIPOHMIQKTRKKRKYFHHNQLDWDHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLY 416
 DB 393 LGPIPOHMIQKTRKKRKYFHHNQLDWDHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLY 452
 QY 417 RRMLEYDPTORTTLDLALQHPFDLLKKX 445
 DB 453 RRMLEYDPTORTTLDLALQHPFDLLKKX 481
 RESULT 3
 AAM49914
 ID AAM49914 standard; Protein; 479 AA.
 XX
 AC AAM49914;
 XX
 DT 20-JUN-1998 (first entry)
 XX
 DB Mouse CLK serine/threonine kinase mCLK4.
 XX
 KW mCLK4; CLK; serine/threonine kinase; protein kinase; LAMMER kinase;

KW signal transduction; cancer; contraceptive; mouse; therapy;
 KW diagnosis.
 XX
 OS Mus musculus.
 XX
 FH Key
 FT Domain
 FT Domain
 FT Domain
 FT Peptide
 FT
 XX
 XX WO9748723-A2.
 XX
 PD 24-DEC-1997.
 XX
 PD 17-JUN-1997; 97WO-IB00946.
 XX
 PF 19-DEC-1996; 96US-0034286.
 PR 17-JUN-1996; 96US-0019629.
 PR 09-AUG-1996; 96US-0023485.
 PR 13-NOV-1996; 96US-0030860.
 PR 15-NOV-1996; 96US-0030964.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Aoki N, Chen Z, Kharitonov AI, Kim YW, Nayler O;
 PI Ullrich A, Wang HY;
 XX
 DR WPI: 1998-120302/11.
 XX
 PT New phosphatase and kinase enzyme(s) - useful in the diagnosis and
 PT treatment of signal transduction disorders
 XX
 PS Claim 11; Fig 4; 138pp; English.
 XX
 CC This polypeptide comprises novel mouse CLK serine/threonine kinase
 CC mCLK4, from the CLK serine/threonine kinase family of proteins that
 CC regulate RNA splicing in cells. mCLK4 cDNA was cloned from a mouse
 CC embryo 11.5 p.c. 12AP cDNA library. The invention discloses the
 CC discovery of novel protein kinases mCLK2 (see AAM4912), mCLK3 and
 CC mCLK4 (see AAM4914) of mol.wt. 59.9, 58.5 and 57.2 kDa,
 CC respectively, as well as other novel proteins (see AAM4906-10)
 CC involved in cellular signal transduction, and provides vectors,
 CC host cells, purified recombinant proteins, methods for identifying
 CC compounds that activate or inhibit the novel proteins, as well as
 CC methods for the diagnosis and treatment of diseases associated with
 CC the novel proteins. Overexpression of CLK serine/threonine kinases
 CC has been implicated in certain types of cancer. Compounds that
 CC inhibit their catalytic activity or disrupt their interactions
 CC with natural binding partners may act as anti-cancer therapeutics.
 CC mCLK related molecules and compounds may also be useful as male
 CC contraceptives.
 XX
 SQ Sequence 479 AA;
 Query Match 90.4%; Score 2177.5; DB 19; Length 479;
 Best Local Similarity 91.1%; Pred. No. 3.4e-199;
 Matches 409; Conservative 9; Mismatches 18; Indels 13; Gaps 4;
 QY 8 SH-SVEEDTH-----PSHYLEARSINERDYRDRRYVDEYRNDYCEGYVPRHYRD 56
 DB 33 SHSSTQENRHKCPHHQFKESDCHYLEARCLNERDYRD-RYIDEYRNDYCEGYVPRHYRD 91
 QY 57 IESGRIHCSKSSVSRSSSPKRNKRHCSSHQSSKSHRRKRSISIEDDEGHILICQSG 116
 DB 92 VESTYRIHCSKSSVSRSSSPKRNKRHCSSHQSSKSHRRKRSISIEDDEGHILICQSG 151
 QY 117 DVLARAYEIVDTLGGAGFGKVECTIDHGMDCGHVAVKIVKNGYRFAARSTIOVLEHNL 176
 DB 152 DVLARAYEIVDTLGGAGFGKVECTIDHGMDCGHVAVKIVKNGYRFAARSTIOVLEHNL 211

OY 177 STDNSVRCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQMAVQICQ 236
 DB 212 STDNSVRCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQMAVQICQ 271
 OY 237 SINFLHNKLTHTDLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 296
 DB 212 SINFLHNKLTHTDLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 330
 OY 297 HNSTLSTRHRAPEVILALGMSQPCDVWSIGCIIIEYVLGFTVPGTHDSKEHLAMERI 356
 DB 331 HNSTLSTRHRAPEVILALGMSQPCDVWSIGCIIIEYVLGFTVPGTHDSKEHLAMERI 390
 OY 357 LGPIPDHMIQKTRKRYFHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416
 DB 391 LGPIPDHMIQKTRKRYFHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 450
 OY 417 RRMLEYDPTORITLDEALQHPFDLLKKK 445
 DB 451 RRMLEYDPTORITLDEALQHPFDLLKKK 479

RESULT 4
 AAB65648
 ID AAB65648 standard; Protein: 499 AA.
 AC AAB65648;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Novel protein kinase, SEQ ID NO: 175.
 XX
 KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
 KW immunosuppressive; carbant; renal; antiinflammatory; antiaesthetic;
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
 XX
 OS Homo sapiens.
 FN WC200073469-A2.
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000WO-US14842.
 XX
 PR 28-MAY-1999; 99US-0136503.
 XX
 PA (SUGF-) SUGEN INC.
 XX
 PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
 DR WPI; 2001-032161/04.
 DR N-PSDB; AAF44675.
 XX
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 XX
 PS Claim 10; Fig 1; 310pp; English.
 XX
 CC The present sequence is a novel protein kinase. The novel protein kinases
 CC and the nucleic acids that encode them may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,

CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.
 XX
 SQ Sequence 499 AA;
 Query Match 88.3%; Score 2127.5; DB 22; Length 499;
 Best Local Similarity 89.3%; Pred. No. 2.2e-194;
 Matches 401; Conservative 2; Mismatches 5; Indels 41; Gaps 3;
 OY 8 SH-SVEBDTH-----PSHYLEARSLNERDYRDRRYVDEYNDYCEGYVPRHYRD 56
 DB 81 SHSTQENRHCPRHGFKESDCHYLEANSLNERDYRDRRYVDEYNDYCEGYVPRHYRD 140
 OY 57 IESGRIHCSKSSVRSRRSPKRRNRNCSHQRSKRRKRSRIEDDEGHLICQSG 116
 DB 141 IESGRIHCSKSSVRSRRSPKRRNRNCSHQRS----- 176
 OY 117 DVLRARVEIVDTLGEAGFAGKVVCEIDHGMGMHVAVKIVKRVGRYRARSSEIOVLEHLN 176
 DB 177 -----XETVDTLGEAGFAGKVVCEIDHGMGMHVAVKIVKRVGRYRARSSEIOVLEHLN 230
 OY 177 STDNSVRCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQMAVQICQ 236
 DB 231 STDNSVRCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQMAVQICQ 290
 OY 237 SINFLHNKLTHTDLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 296
 DB 291 SINFLHNKLTHTDLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 350
 OY 297 HNSTLSTRHRAPEVILALGMSQPCDVWSIGCIIIEYVLGFTVPGTHDSKEHLAMERI 356
 DB 351 HNSTLSTRHRAPEVILALGMSQPCDVWSIGCIIIEYVLGFTVPGTHDSKEHLAMERI 410
 OY 357 LGPIPDHMIQKTRKRYFHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416
 DB 411 LGPIPDHMIQKTRKRYFHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 470
 OY 417 RRMLEYDPTORITLDEALQHPFDLLKKK 445
 DB 471 RRMLEYDPTORITLDEALQHPFDLLKKK 499

RESULT 5
 AAY27054
 ID AAY27054 standard; Protein: 451 AA.
 XX
 AC AAY27054;
 XX
 DT 08-OCT-1999 (first entry)
 XX
 DE Human protein kinase (HPKM)-3 (clone ID 339963).
 XX
 KW Human protein kinase molecule; HPKM; human; protein kinase;
 KW phosphate group; cancer; immune disorder.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Modified-site 182
 FT Domain 135..159
 FT /note= "potential signature sequence for protein
 FT kinase catalytic domain"
 FT Domain 252..264
 FT /note= "potential signature sequence for protein
 FT kinase catalytic domain"
 FT Modified-site 4
 FT /note= "potential phosphorylation site"
 FT Modified-site 22
 FT /note= "potential phosphorylation site"

50 Sequence 484 AA;

Best Local Similarity 82.1%; Pred.No. 21e-171;
Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

78 KKK-KNRHCSSH-QSRSKSHRRKRSRSIEDDEEGHLICQSGDVLRARYEIVDTLGEAFC 135

150 AVECLDHGMGMHVAVKI VKNVGRYKEARSEIQVLEHLNSTDPNSVFRVCQMLEWFDH 195

255

[illegible]

5

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525</

474 HPEEDITYR 483

Matches	347;	Conservative	33;	Mismatches	51;	Indels	7;	Gaps	5;
---------	------	--------------	-----	------------	-----	--------	----	------	----

9 H S V E E D H P S H Y L E A R S L N E R D Y R D R Y V D E T R N D Y C E G I V P R H T H R D I E G Y R I H S K S 60

```

Db      48 HSKTTD---SYLSESRSEINEKAYHSRRYVDEYNDYM-GYEPGHPIGEGPSGRYOMSSKS 103
QY      69 SVASRRSPKRR-RNR-HCSSHOSRSKSHRRKRSRS-IEDDEGHILIQSGDVLMARYEIV 126
Db      104 SGSSGRSSYSKSHRSRHHTSQHSHDGHSHRRKRSRSVDEDEGHILIQSGDVLSARYEIV 163
QY      127 DTGEGAFGKRVNEDIDHGMGMHAYVIVKRVGRRARASEIOVLEHLNSTDPNSVFC 186
Db      164 DTGEGAFGKRVNEDIDHGMGMHAYVIVKRVGRRARASEIOVLEHLNSTDPNSVFC 223
QY      187 VQMLEWFDHGHVCIYVELLGLSTYDFIKENSFLPFOIDIRQMAVOICOSINFLHNKL 246
Db      224 VQMLEWFEHRGHICIVVELLGLSTYDFIKENSFLPFOIDIRQMAVOICOSINFLHNKL 283
QY      247 THTDLPENILFYKSDIVVYKNSKMRDERTLNKTDIKVDPGSAFYDDEHSHSTVSTRH 306
Db      284 THTDLPENILFYKSDIVVYKNSKMRDERTLNKTDIKVDPGSAFYDDEHSHSTVSTRH 342
QY      307 YRAPEVILALGWSQPCDVMSIGCILLEYLIGFTVFOHDSKEHLAMMERILGLPIQMTIQ 366
Db      343 YRAPEVILALGWSQPCDVMSIGCILLEYLIGFTVFOHDSKEHLAMMERILGLPIQMTIQ 402
QY      367 KTRKRYFFHNOIDMDHSHSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLBYDPTQ 426
Db      403 KTRKRYFFHNDRLDMDHSHSAGRYVRRCKPLKEFMLSQDAHEHFLFDLVGKILLETDPK 462
QY      427 RITLDEALQHPFDLLKK 444
Db      463 RITLKEALKHPFFYPLKK 480

RESULT 8
AAB33778
ID      AAB33778 standard; Protein; 283 AA.
XX
AC      AAB33778;
XX
DT      02-FEB-2001 (first entry)
XX
DE      Human secreted protein BLAST search protein SEQ ID NO: 122.
XX
KW      Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW      antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW      vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW      cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein.
XX
OS      Homo sapiens.
XX
PN      WO200056753-A1.
XX
PD      28-SEP-2000.
XX
PF      16-MAR-2000; 2000WO-US06765.
XX
PR      23-MAR-1999; 99US-0126051.
XX
PR      10-DEC-1999; 99US-0169906.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Rosen CA, Ruben SM, Komatsoulis G;
XX
DR      WPI, 2000-594570/56.
XX
PT      Nucleic acid molecules encoding human secreted proteins, used in
XX
PS      preventing, treating or ameliorating a disorder -
XX
PS      Disclosure, Page 400-401; 410pp; English.
XX
CC      The invention relates to the isolation of genes AAC59277-C59325 encoding
CC      the human secreted proteins AAB33718-B33764. The sequence is used as a
CC      query sequence for doing BLASTX searches to identify homologous
CC      sequences. The genes and proteins are useful for preventing,

```

```

CC      ameliorating or treating medical conditions, e.g. by protein or gene
CC      therapy. The genes are isolated from a range of human tissues disclosed
CC      in the specification. The nucleic acids, proteins, antibodies and
CC      (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC      (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC      adrenal gland, bone, bone marrow, breast, gastrointestinal
CC      tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
CC      disease, allergies, autoimmune haemolytic anaemia, autoimmune
CC      thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
CC      rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
CC      disorders such as myocardial ischaemia; (d) wound healing; (e)
CC      neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC      infectious diseases such as viral, bacterial, fungal and parasitic
CC      infections.
XX
SQ      Sequence 283 AA;
XX
Query Match 63.7%; Score 1535; DB 21; Length 283;
Best Local Similarity 100.0%; Pred. No. 4, 2e-138;
Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      163 EAARSEIOVLEHINSTDPNSVFCVQMLEWFDHGHVCIYVELLGLSTYDFIKENSFLPF 222
Db      1 EAARSEIOVLEHINSTDPNSVFCVQMLEWFDHGHVCIYVELLGLSTYDFIKENSFLPF 60
QY      223 QIDHIRQMAVOICOSINFLHNKLTHTDLPENILFYKSDIVVYKNSKMRDERTLNKTD 282
Db      61 QIDHIRQMAVOICOSINFLHNKLTHTDLPENILFYKSDIVVYKNSKMRDERTLNKTD 120
QY      283 IKVDPGSAFYDDEHSHSTVSTRHRAPEVILALGWSQPCDVMSIGCILLEYLIGFTVFO 342
Db      121 IKVDPGSAFYDDEHSHSTVSTRHRAPEVILALGWSQPCDVMSIGCILLEYLIGFTVFO 180
QY      343 THDSKEHLAMMERILGLPIQMTIQTRKRYFFHNOIDMDHSHSAGRYVRRCKPLKEFM 402
Db      181 THDSKEHLAMMERILGLPIQMTIQTRKRYFFHNOIDMDHSHSAGRYVRRCKPLKEFM 240
QY      403 LCHDEHEKLFDLVRMLEYDPTQRTTLDEALQHPFDLLKK 445
Db      241 LCHDEHEKLFDLVRMLEYDPTQRTTLDEALQHPFDLLKK 283

RESULT 9
AAB58831
ID      AAB58831 standard; Protein; 350 AA.
XX
AC      AAB58831;
XX
DT      27-MAR-2001 (first entry)
XX
DE      Breast and ovarian cancer associated antigen protein sequence SEQ ID 539.
XX
KW      Human; breast cancer; ovarian cancer; cytosstatic; immunosuppressive;
KW      nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW      antidiabetic; antiinflammatory; anticancer; vulnerrary; anticonvulsant;
KW      antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW      Addison's disease; allergy; autoimmune haemolytic anaemia;
KW      autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW      multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW      cardiovascular disorder; wound healing; neurological disease.
XX
OS      Homo sapiens.
XX
PN      WO200055173-A1.
XX
PD      21-SEP-2000.
XX
PF      08-MAR-2000; 2000WO-US05881.
XX
PR      12-MAR-1999; 99US-0124270.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX

```

PI Rosen CA, Ruben SM;
XX WPI; 2000-611515/58.
DR N-PSDB; AAF21734.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 11: Page 970-971; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antibacterial; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
CC antifungal; antiparasitic and cardiac activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular diseases such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 350 AA;

Query Match 63.3%; Score 1526; DB 21; Length 350;
Best Local Similarity 86.9%; Pred. No. 4.1e-137;
Matches 279; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 124 EIVDTLGGAGFGKVEECIDHGMGNVAVKIVNNGRYREARSEIOVLEHINSTDPNSV 183
DB 28 EIVDTLGGAGFGKVEECIDHGMGNVAVKIVNNGRYREARSEIOVLEHINSTDPNST 87
QY 184 FRCVOMLEMPDHGHVGVCFEELGISTYDFIKENSFLPQIHNQMAVOICQSNFLNH 243
DB 88 FRCVOMLEMPDHGHVGVCFEELGISTYDFIKENSFLPQIHNQMAVOICQSNFLNH 147
QY 244 NKLATLTKPEKENTLFEVKSDDYVVKYNSKMKDERTLKNTDIKVDGSGATYDEHSTVS 303
DB 148 NKLATLTKPEKENTLFEVKSDDYVVKYNSKMKDERTLKNTDIKVDGSGATYDEHSTVS 207
QY 304 TRHYRAPVILALGMSQPCDWSIGCIIIEYVLGFTVQTHDSKHEHLLAMERILGPIPOH 363
DB 208 TRHYRAPVILALGMSQPCDWSIGCIIIEYVLGFTVQTHDSKHEHLLAMERILGPIPOH 267
QY 364 MIOKRRKRYFHHNOLDMDDEHSSAGRYVRRCKPKLKEPMLCHDEHEKFLDVRMLEYD 423
DB 268 MIOKRRKRYFHHNOLDMDDEHSSAGRYVRRCKPKLKEPMLCHDEHEKFLDVRMLEYD 327
QY 424 FTORITLDEALQHPFDLLKK 444
DB 328 FAKRITLREALKHPFDLLKK 348

RESULT 10

AAB33777 ID AAB33777 standard; Protein; 283 AA.

XX AAB33777;

XX 02-FEB-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 121.

KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO200056753-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06765.

XX 23-MAR-1999; 99US-0126051.

XX 10-DEC-1999; 99US-0169906.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594570/56.

XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder -

XX Disclosure; Page 399-400; 410pp; English.

CC The invention relates to the isolation of genes AAF59277-C59325 encoding
CC the human secreted proteins AAB33718-B33764. The sequence is a search
CC result from a BLASTX homology search. The genes and proteins are useful
CC for preventing, ameliorating or treating medical conditions, e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.

XX Sequence 283 AA;

Query Match 62.9%; Score 1517; DB 21; Length 283;
Best Local Similarity 98.6%; Pred. No. 2.2e-136;
Matches 279; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 163 EAAREIOVLEHINSTDPNSVFRQVOMLEMPDHGHVGVCFEELGISTYDFIKENSFLPF 222
DB 1 EAAREIOVLEHINSTDPNSVFRQVOMLEMPDHGHVGVCFEELGISTYDFIKENSFLPF 60
QY 223 QIDHTRQMAVOICQSNFLHNKLTHTDLPKENTLFEVKSDDYVVKYNSKMKDERTLKNTD 282
DB 61 QIDHTRQMAVOICQSNFLHNKLTHTDLPKENTLFEVKSDDYVVKYNSKMKDERTLKNTD 120
QY 283 IKVVDGSGATYDEHSTVS TRHYRAPVILALGMSQPCDWSIGCIIIEYVLGFTVQ 342
DB 121 IKVVDGSGATYDEHSTVS TRHYRAPVILALGMSQPCDWSIGCIIIEYVLGFTVQ 180
QY 343 THDSKEHLLAMERILGPIPOH MIOKRRKRYFHHNOLDMDDEHSSAGRYVRRCKPKLKEP 402
DB 181 THDSKEHLLAMERILGPIPOH MIOKRRKRYFHHNOLDMDDEHSSAGRYVRRCKPKLKEP 240
QY 403 LCHDEHEKFLDVRMLEYDPTQITLDEALQHPFDLLKK 445
DB 241 LCHDEHEKFLDVRMLEYDPTQITLDEALQHPFDLLKK 283

CC This polypeptide comprises novel mouse CLK serine/threonine kinase
CC mCLK2, from the CLK serine/threonine kinase family of proteins that
CC regulate RNA splicing in cells. mCLK2 cDNA was cloned from a mouse
CC embryo 11.5 p.c. 12AP cDNA library. The invention discloses the
CC discovery of novel protein kinases mCLK2, mCLK3 (see AAM49912) and
CC mCLK4 (see AAM49914) of mol.wt. 59.9, 58.5 and 57.2 kDa,
CC respectively, as well as other novel proteins (see AAM49906-10)
CC involved in cellular signal transduction, and provides vectors,
CC host cells, purified recombinant proteins, methods for identifying
CC compounds that activate or inhibit the novel proteins, as well as
CC methods for the diagnosis and treatment of diseases associated with
CC the novel proteins. Overexpression of CLK serine/threonine kinases
CC has been implicated in certain types of cancer. Compounds that
CC inhibit their catalytic activity or disrupt their interactions
CC with natural binding partners may act as anti-cancer therapeutics.
CC mCLK related molecules and compounds may also be useful as male
CC contraceptives.

XX Sequence 497 AA;

Query Match 55.2%; Score 1311.5; DB 19; Length 497;
Best Local Similarity 59.0%; Pred. No. 2.7e-118;
Matches 259; Conservative 62; Mismatches 105; Indels 13; Gaps 8;

QY 13 EPTFSSHYLEARS-----LNERDYDRRYVDEY-RNDYCEGYVPRHYRDISSGYRIHCS 66
DB 47 EDSR--YHVRSSSDYDHSRRLYDRYCSGYRNDYGRDGEAYVDTDFQSYEYHRE 104
QY 67 KSVASRRSS-PKRENHCSHQSRSKSHRRKRSSTIDDEGHILCOGQVLAARYI 125
DB 105 NSSYSSQSSRRKRRRRRRRTFSRSSHSSRAKVSVDDEGHILYHVGWMLDERYEI 164
QY 126 VDTLEGAFGKVEECIDHGMGMHVAVKIVKVGREARAEIQYLEHINSTDPNSVPR 185
DB 165 VSTLEGGSGRNVGICIDRRVGRNRYLKIKNVEKKEARLEINVEKINERKDKNTL 224
QY 186 CYQMLEPDPHGHVCIPELLGLSTYDPDKENSFLPFOJDHTRQMAVOICQINFLHNK 245
DB 225 CYQMDWDPYHGHMCISPELLGLSTPDKNNYLPYPIHQVRHMAFOLCAVKFLHDNK 284
QY 246 LHTDTLKPENILFVSDYVVKYNSMKKRDERT-LKNTDIXKVDPSGATDDEHHSITV 304
DB 285 LHTDTLKPENILFVNSDYELT-NPLEKRDERTSVSSTAR-VDPGATFDDHHSITLST 342
QY 305 RHYRAPEVILLGWSQPCDWSIGCILLEYVLGFTVFCOTDSKEHLAMMERILGPIPOHM 364
DB 343 RHYRAPEVILLGWSQPCDWSIGCIFLEYVLGFTVPHNSRHLAMMERILGVPSPRM 402
QY 365 IOKTKRKYFHNOLDMDHBSAGRYVRRCKPLKEFMLCHDEEHEKFLDYLRMLEYDP 424
DB 403 IRTKRQKYFYGRDMDENTSAGRYVRENCKPLRKY-LTSEADHQLFDLIENMLEYEP 461
QY 425 TORITLDEALQHPFFDLK 443
DB 462 AKRLTLGHALQHPFACLR 480

RESULT 13
ABG23354
ID ABG23354 standard; Protein; 490 AA.
XX
AC ABG23354;

XX 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23345.
XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX Food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
XX MO200175067-AA.

XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PA
XX (HSE-) HSEQ INC.
PI
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
XX N-PSDB; AAS87541.
PT
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

Claim 20; SEQ ID No 53713; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 490 AA;

Query Match 54.5%; Score 1314.5; DB 22; Length 490;
Best Local Similarity 59.8%; Pred. No. 1.1e-116;
Matches 251; Conservative 59; Mismatches 89; Indels 21; Gaps 7;

QY 34 RRYVDE--YRNDYCEGYVPR-HYHRDISSGYRIHCSKSVRSRSPKRX-----NR 83
DB 64 RRYHSSERGSRSYEHNRSRKHQRRSRS-----WSSSDSDTRR---RRRDSYHVR 114
QY 84 HCSSHQSRSKS-HRRKRSRSIDDEEGHILCOGQVLAARYI VDTLGEAGFQVVECID 142
DB 115 RCSRFTSSRSSQSHSSRKAKSVEDDTEGHILYHVGWMLDERYIVGLTGTFGRVVOQVD 174
QY 143 HGMGMHVAVKIVKVGREARAEIQYLEHINSTDPNSVRCYQMLEPDPHGHVCI 202
DB 175 HRRRGARVALKTIKNVEYKEARLEIVLEKINERKDPQKNL-CYQMDWDPYHGHMCIS 233
QY 203 FELLGLSTYDPIKENSFLPFOJDHTRQMAVOICQINFLHNKLTHTDILKPENILFVKSD 262
DB 234 LELLGLSTPDLKXNNHLPYPIHQVRHMAFOLCAVKFLHDNKLHTDILKPENILFVNSD 293
QY 263 YVVKYNSCMKGRDEPRLKNTDIXKVDPSGATYDDEHHSITVSTRHRAPEVILLGWSQPC 322
DB 294 YELTYNLKKKHRSRVSKTAVKVGDFGSATPDEHHSITVSTRHRAPEVILLGWSQPC 353
QY 323 DVMSIGCLILEYVLGFTVFCOTDSKEHLAMMERILGPIPOHMIOKTKRKYFHNOLDMD 382
DB 354 DVMSIGCLILEYVGLFTVFCOTDSKEHLAMMERILGPIPSIRTKRKQKIFYRGALDWD 413

QY 383 EHSAGRYVRRCPELKEFMLCHDEHEKLDLVRLMLEYDPTORITLDEALQHPFDLL 442
 DB 414 ENTSAGRYVRBCKPLRQVLTSEAEEDHQFDLIESMLEYEPARQLTLGALQHPFPRSL 473

RESULT 14
 ID AAB85506 standard; protein; 431 AA.
 AC AAB85506;
 DT 25-SEP-2001 (first entry)
 XX
 DE Human protein kinase SGK090.
 XX
 KW Protein kinase; enzyme; cytosolic; nootropic; neuroprotective; human;
 KW antiParkinsonian; virucide; antibacterial; antifungal; antimigraine;
 KW analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;
 KW antisoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;
 KW osteopathic; thrombolytic; antiarteriosclerotic; antiastrumatic;
 KW vasotropic; antidiabetic; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 27 /note= "encoded by CAN"
 FT Misc-difference 28 /note= "encoded by NAG"
 FT
 PN MO200155356-A2.
 PD 02-AUG-2001.
 PF 25-JAN-2001; 2001WO-US02337.
 PR 25-JAN-2000; 2000US-0178078.
 PR 31-JAN-2000; 2000US-0179364.
 PR 17-FEB-2000; 2000US-0183173.
 PR 17-MAR-2000; 2000US-0190162.
 PR 29-MAR-2000; 2000US-0193404.
 PR 13-NOV-2000; 2000US-0247013.
 XX
 PA (SUGEN-) SUGEN INC.
 XX
 PI Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;
 XX
 DR WPI; 2001-476202/51.
 DR N-PSDB; AAB46906.
 XX
 PT Kinase polypeptides useful for treating cancers, Alzheimer's disease,
 PT viral infections, diabetes, obesity, organ transplant rejection and
 PT rheumatoid arthritis -
 XX
 PS Claim 7; Page 216; 218bp; English.
 XX
 CC The invention provides human protein kinases and protein kinase-like
 CC enzymes and polynucleotides encoding the polypeptides. The kinase
 CC polypeptides and their modulators are useful for treating a disease or
 CC disorder such as cancer, immune-related diseases, cardiovascular disease,
 CC brain or neuronal-associated disease and metabolic disorders, including
 CC cancers of tissues, cancers of hematopoietic origin, diseases of the
 CC central nervous system, diseases of the peripheral nervous system,
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
 CC lateral sclerosis, viral infections, infections caused by prions,
 CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,
 CC mood disorders, attention disorders, cognitive disorders, hypotension,
 CC hypertension, psychotic disorders, neurological disorders, dyskinesias,
 CC metabolic disorders, and organ transplant rejection. They are also useful
 CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,
 CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic
 CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders

CC such as diabetes, obesity, cardiovascular diseases such as reperfusion
 CC injury, coronary thrombosis, clotting disorders and atherosclerosis,
 CC ocular diseases such as glaucoma, retinopathy and macular degeneration,
 CC psychiatric and neurological disorders such as anxiety, schizophrenia,
 CC dementia, manic depression, etc. The polynucleotides are useful in gene
 CC therapy techniques to treat the above mentioned disorders. Sequences
 CC AAB85491-85522 represent the human protein kinases of the invention.
 XX
 SQ Sequence 431 AA;
 Query Match 54.4%; Score 1310.5; DB 22; Length 431;
 Best Local Similarity 59.2%; Pred. No 2.2e-116;
 Matches 248; Conservative 61; Mismatches 91; Indels 19; Gaps 6;

QY 34 RRYVDE--YRNDYCEGVPRHYRHDIESGYRIHCKSSVSRSSPKRRK-----NRH 84
 DB 5 RRYHSSBRGSRGSGYCEHYRGRKXXQRSRSM---SSSDRTR-----RRRDSYHVR 56
 QY 85 CSHQSSKSKS-HRRKRSRSLEDDEBGLICQSGDVLPRRIEIVDTLGBAGFYVECTIDH 143
 DB 57 CSRTFSRSSQSHSRKAKSVYEDDTGHLIYVGDWLQERYEIVSTLKGTFGRVVCVDH 116
 QY 144 GMDQMHVAVKIYVGVGRYREARSEIQVLEHLNSTDPNSVPRCYQMLEWPHGHVCI 203
 DB 117 RRRGARVALKIIQVKEKRAALEIKYLEKINCKDGRKL-CYQMDWPDYHGMICSL 175
 QY 204 ELIGLSTYDFIKENSLFQIDIRIOMAYOICQSLNPLHNKLTHTDLKENLILFYKSDY 263
 DB 176 ELIGLSTFDPLKDNHLPYPIHQVHMAQLOCAVKEFLHDKLTHTDLKENLILFVNSDY 235
 QY 264 VVKYNSKMKRDEFTLKNITDKVYDFGSATYDDDEHSHLTVSTRYRAEVLALGMSQPCD 323
 DB 236 ELTPNLEKRRHESVSTARVDFGSAFDHEHSHVISTRYRAEVLILEGMSQPCD 295
 QY 324 VMSIGCLILEYVGFVQTHDSKENLHMMERITLGPQMIQKTRKRYFHHNQLMDE 383
 DB 296 VMSIGCLIFEYVGFVFLFQTHDRQHLATMERILGPSMKIKTRKQKFFRRLDWE 355
 QY 384 HSSAGRYVRRCPELKEFMLCHDEHEKLDLVRLMLEYDPTORITLDEALQHPFDLL 442
 DB 356 NTSAGRYVRBCKPLRQVLTSEAEEDHQFDLIESMLEYEPARQLTLGALQHPFPRSL 414

RESULT 15
 ID AAB49913 standard; Protein; 485 AA.
 AC AAB49913;
 DT 20-JUL-1998 (first entry)
 XX
 DE Mouse CLK serine/threonine kinase mCLK3.
 XX
 KW mCLK3; CLK; serine/threonine kinase; protein kinase; LAMMER kinase;
 KW signal transduction; cancer; contraceptive; mouse; therapy;
 KW diagnosis.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT Domain 23..41
 FT Domain /note= "nuclear localisation domain"
 FT Domain 151..469
 FT Peptide /note= "catalytic domain"
 FT Peptide 377..382
 FT Peptide /note= "LAMMER motif"
 PN MO9748723-A2.
 PD 24-DEC-1997.
 XX
 PF 17-JUN-1997; 97WO-IB00946.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:16:24 ; Search time 21 Seconds
(without alignments)
896.587 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410
Sequence: 1 MCIPLEASHSVEDTHPSHY.....QRTLDALQHPDILKKK 445

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2410	100.0	445	4	US-09-810-671-2 Sequence 2, Appl1
2	2312	95.9	427	4	US-09-810-671-4 Sequence 4, Appl1
3	2117.5	87.9	451	2	US-09-016-000-3 Sequence 3, Appl1
4	1887	78.3	429	4	US-09-810-671-5 Sequence 5, Appl1
5	1887	78.3	484	4	US-09-457-040B-12 Sequence 12, Appl1
6	1417	58.8	499	4	US-09-457-040B-13 Sequence 13, Appl1
7	1222	50.7	517	4	US-09-457-040B-14 Sequence 14, Appl1
8	695	28.8	736	4	US-09-457-040B-26 Sequence 26, Appl1
9	550.5	22.8	527	4	US-09-659-166-2 Sequence 2, Appl1
10	546.5	22.7	528	2	US-08-802-466-2 Sequence 2, Appl1
11	546.5	22.7	528	3	US-09-350-484-2 Sequence 2, Appl1
12	544.5	22.6	568	2	US-08-835-170-4 Sequence 4, Appl1
13	544.5	22.6	568	3	US-09-359-257-4 Sequence 4, Appl1
14	544.5	22.6	568	4	US-09-371-674-4 Sequence 4, Appl1
15	544.5	22.6	588	2	US-08-835-170-4 Sequence 2, Appl1
16	544.5	22.6	588	3	US-09-359-257-2 Sequence 2, Appl1
17	544.5	22.6	588	4	US-09-371-674-2 Sequence 2, Appl1
18	523	21.7	508	2	US-08-818-024-3 Sequence 3, Appl1
19	523	21.7	508	3	US-09-334-775A-3 Sequence 3, Appl1
20	517	21.5	620	3	US-09-126-646-2 Sequence 2, Appl1
21	517	21.5	620	4	US-09-421-491-2 Sequence 2, Appl1
22	505	21.0	539	2	US-08-818-024-4 Sequence 4, Appl1
23	505	21.0	539	3	US-09-334-775A-4 Sequence 4, Appl1
24	505	21.0	539	4	US-08-789-275-6 Sequence 6, Appl1
25	478.5	19.9	557	3	US-09-027-064-2 Sequence 2, Appl1
26	478.5	19.9	557	4	US-09-271-815-2 Sequence 2, Appl1
27	472.5	19.6	763	2	US-08-677-862-2 Sequence 2, Appl1

28	472.5	19.6	763	2	US-09-252-571-2 Sequence 2, Appl1
29	472.5	19.6	763	3	US-09-434-065-2 Sequence 2, Appl1
30	472.5	19.6	763	3	US-08-789-275-4 Sequence 4, Appl1
31	472.5	19.6	763	3	US-08-789-275-5 Sequence 5, Appl1
32	419	17.4	1209	4	US-09-749-588-4 Sequence 4, Appl1
33	407	16.9	1170	4	US-09-749-588-2 Sequence 2, Appl1
34	392	16.3	1087	1	US-08-264-002-5 Sequence 5, Appl1
35	359.5	14.9	376	2	US-08-264-002-2 Sequence 2, Appl1
36	359.5	14.9	376	3	US-08-818-024-1 Sequence 1, Appl1
37	343	14.2	353	1	US-09-334-775A-1 Sequence 1, Appl1
38	343	14.2	353	2	US-08-176-620A-14 Sequence 14, Appl1
39	332.5	13.8	607	4	US-08-461-985-14 Sequence 14, Appl1
40	331.5	13.8	655	1	US-09-417-197-47 Sequence 47, Appl1
41	328	13.6	379	4	US-08-264-002-2 Sequence 2, Appl1
42	326.5	13.5	360	1	US-09-457-040B-36 Sequence 36, Appl1
43	326.5	13.5	360	1	US-08-674-612-4 Sequence 4, Appl1
44	326.5	13.5	360	1	US-08-469-421-12 Sequence 12, Appl1
45	326.5	13.5	360	1	US-08-469-421-14 Sequence 14, Appl1
					US-08-250-975-12 Sequence 12, Appl1

ALIGNMENTS

RESULT 1
US-09-810-671-2
; Sequence 2, Application US/09810671
; Patent No. 6455291
; GENERAL INFORMATION:
; APPLICANT: YAN Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: C1000758
; CURRENT APPLICATION NUMBER: US/09/810,671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ. ID NOS: 5
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ. ID NO 2
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Human
; US-09-810-671-2

Query Match 100.0%; Score 2410; DB 4; Length 445;
Best Local Similarity 100.0%; Pred. No. 4, 1e-218;
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MCIPLEASHSVEDTHPSHYLEARSINERDYRDYVDEYNDYCEGYVRHRYDI	ESG 60
DB	1	MCIPLEASHSVEDTHPSHYLEARSINERDYRDYVDEYNDYCEGYVRHRYDI	ESG 60
QY	61	YRHCSSVRSRSPKRRKRRHSSHSRSHRRKRSIEDDEEGLICQSGVLR	120
DB	61	YRHCSSVRSRSPKRRKRRHSSHSRSHRRKRSIEDDEEGLICQSGVLR	120
QY	121	ARVEIVDTLGEAGFGKVVCEIDHGMDSHVAIVKVVGRYRAARSEI	180
DB	121	ARVEIVDTLGEAGFGKVVCEIDHGMDSHVAIVKVVGRYRAARSEI	180
QY	181	NSYFRVCQMLEWDDHGHVCIYVLELGLSTYPIKNSFLPQIDHIR	240
DB	181	NSYFRVCQMLEWDDHGHVCIYVLELGLSTYPIKNSFLPQIDHIR	240
QY	241	LHNKLTHTDLPKENTLFPKSDPVVYKSKMRDETLKNTDKVVD	300
DB	241	LHNKLTHTDLPKENTLFPKSDPVVYKSKMRDETLKNTDKVVD	300
QY	301	LVSTRYRAPAVIALGWSQPCDWSIGCILLEYVIGFTVFTQTHS	360
DB	301	LVSTRYRAPAVIALGWSQPCDWSIGCILLEYVIGFTVFTQTHS	360
QY	361	POHMTOKTKKRYFFHNNQJDMDEHSAGYVRRRCPLKEPM	420
DB	361	POHMTOKTKKRYFFHNNQJDMDEHSAGYVRRRCPLKEPM	420

Db 361 PQHICKTRKRYFHNNQJLMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPJLVRML 420
 QY 421 EYDPTQRIITIDEALQHPFDLLKKK 445
 Db 421 EYDPTQRIITIDEALQHPFDLLKKK 445

RESULT 2
 US-09-810-671-4
 ; Sequence 4, Application US/09810671
 ; Patent No. 6455291
 ; GENERAL INFORMATION:
 ; APPLICANT: YAN, Chunhua et al
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: CLO00758
 ; CURRENT APPLICATION NUMBER: US/09/810,671
 ; CURRENT FILING DATE: 2001-06-08
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 427
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-810-671-4

Query Match 95.9%; Score 2312; DB 4; Length 427;
 Best Local Similarity 100.0%; Pred. No. 6.1e-209;
 Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 HYLEARSLNERDYRDYVDEYRNDYCEGVPRHRYHDIHSGYRHCSSKSVSRSSPK 78
 Db 1 HYLEARSLNERDYRDYVDEYRNDYCEGVPRHRYHDIHSGYRHCSSKSVSRSSPK 60
 QY 79 RKNRHCSSHOSKSHRRKRSRSIEDDEBGLICSGDYLARYETIYDPLGEGAGKVV 138
 Db 61 RKNRHCSSHOSKSHRRKRSRSIEDDEBGLICSGDYLARYETIYDPLGEGAGKVV 120
 QY 139 ECIDHGMQGHVAVKIVKNGRYRRAARSEIQVLEHINSTDPNSVFCVOMLEWPDHGH 198
 Db 121 ECIDHGMQGHVAVKIVKNGRYRRAARSEIQVLEHINSTDPNSVFCVOMLEWPDHGH 180
 QY 199 VCIVFELGLSTYDFIKENSFLPFOIDHIRQMAVYQICQSIINFLHNKLTHTDLKPNILF 258
 Db 181 VCIVFELGLSTYDFIKENSFLPFOIDHIRQMAVYQICQSIINFLHNKLTHTDLKPNILF 240
 QY 259 VKSDYVVKYNSKMKRDERTLKNTDIKVDPGSATYDDEHSTLSTRHYRAPEVILALGM 318
 Db 241 VKSDYVVKYNSKMKRDERTLKNTDIKVDPGSATYDDEHSTLSTRHYRAPEVILALGM 300
 QY 319 SGPQVWSIGCLILEYLGFTVQTHDSKEHLAMMERILGPIPOHMIQTRKRYPHNQ 378
 Db 301 SGPQVWSIGCLILEYLGFTVQTHDSKEHLAMMERILGPIPOHMIQTRKRYPHNQ 360
 QY 379 LMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPJLVRMLBYDPTQRIITIDEALQHPF 438
 Db 361 LMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPJLVRMLBYDPTQRIITIDEALQHPF 420
 QY 439 FDLKKK 445
 Db 421 FDLKKK 427

RESULT 3
 US-09-016-000-3
 ; Sequence 3, Application US/09016000
 ; Patent No. 5962232
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Ial, Preeti
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Akerdion, Ingrid E.

APPLICANT: Shah, Purvi
 APPLICANT: Corley, Neil C.
 APPLICANT: Guegler, Karl G.
 TITLE OF INVENTION: PROTEIN KINASE MOLECULES
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/016,000
 FILING DATE: HEREWITH
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0465 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 451 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: NEUTPMT01
 CLONE: 339963
 US-09-016-000-3

Query Match 87.9%; Score 2117.5; DB 2; Length 451;
 Best Local Similarity 88.9%; Pred. No. 1.2e-190;
 Matches 399; Conservative 3; Mismatches 6; Indels 41; Gaps 3;

QY 8 SH-SVEEDTH-----PSHYLEARSLNERDYRDYVDEYRNDYCEGVPRHRYHDI 56
 Db 33 SHSTQENRCHCKPHHQFKESDCHYLEARSLNERDYRDYVDEYRNDYCEGVPRHRYHDI 92
 QY 57 IESGYRHCSSKSVSRSSPKRKNRHCSSHOSKSHRRKRSRSIEDDEBGLICQSG 116
 Db 93 IESGYRHCSSKSVSRSSPKRKNRHCSSHOSKSHRRKRSRSIEDDEBGLICQSG 128
 QY 117 DVLARAYETIYDPLGEGAGKVVETIDHGMQGHVAVKIVKNGRYRRAARSEIQVLEHIN 176
 Db 129 -----MKSVDPLGEGAGKVVETIDHGMQGHVAVKIVKNGRYRRAARSEIQVLEHIN 182
 QY 177 STDPSVTRCVOMLEWPDHGHVCTIVFELGLSTYDFIKENSFLPFOIDHIRQMAVYQICQ 236
 Db 183 STDPSVTRCVOMLEWPDHGHVCTIVFELGLSTYDFIKENSFLPFOIDHIRQMAVYQICQ 242
 QY 237 SINFLHNKLTHTDLKPNILFVKSDYVVKYNSKMKRDERTLKNTDIKVDPGSATYDDE 296
 Db 243 SINFLHNKLTHTDLKPNILFVKSDYVVKYNSKMKRDERTLKNTDIKVDPGSATYDDE 302
 QY 297 HHSTLSTRHYRAPEVILALGMSGPCDVWSIGCLILEYLGFTVQTHDSKEHLAMMERI 356
 Db 303 HHSTLSTRHYRAPEVILALGMSGPCDVWSIGCLILEYLGFTVQTHDSKEHLAMMERI 362
 QY 357 LGPIPOHMIQTRKRYFHNNQJLMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPJLVRML 416

Db 363 LGPIPOHMIQKTRKRYFHHNQLDWDDEHSSAGRYVRRCCKPLKEFMLCHDEHEKLPDLV 422
 QY 417 RRMLEYDPTORITLDEALQHPFDLKK 445
 Db 423 RRMLEYDPTORITLDEALQHPFDLKK 451

RESULT 4
 US-09-810-671-5
 ; Sequence 5, Application US/09810671
 ; Patent No. 6455291
 ; GENERAL INFORMATION:
 ; APPLICANT: YAN, Chunhua et al
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: C1000758
 ; CURRENT APPLICATION NUMBER: US/09/810,671
 ; CURRENT FILING DATE: 2001-06-08
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 429
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-810-671-5

Query Match 78.3%; Score 1887; DB 4; Length 429;
 Best Local Similarity 82.1%; Pred. No. 5,2e-169;
 Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY 18 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 77
 Db 1 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 60
 QY 78 KKR-RNRKSSH-QSRKSHRRKRSRSIIDEDEGHLICQSGDVLARAYEIVDTLGEAGF 135
 Db 61 KSKRIRHSHSTRSHSGSHRRKRTSRVEDDEGHLICQSGDVLARAYEIVDTLGEAGF 120
 QY 136 KVECIDHGMQDMHVAVKYVNGVYRPAARSEIOVLEHNLSTDPNSVFRVCQMLEFSDH 195
 Db 121 KVECIDHKGAGRHVAVKYVNGVYRPAARSEIOVLEHNLSTDPNSVFRVCQMLEFSDH 180
 QY 196 HGVCIIVFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPN 255
 Db 181 HGVCIIVFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPN 240
 QY 256 ILFVKSQDYVVKNSKMKRDERTLKNTDIAKVDPGSATYDDEHSTLVSTRHYRAPEVILA 315
 Db 241 ILFVKSQDYVVKNSKMKRDERTLKNTDIAKVDPGSATYDDEHSTLVSTRHYRAPEVILA 300
 QY 316 LGMSQPCDVMWSIGCIIIEYVIGFTVFOTHDSKEHLAMMERILGPIPOHMIQKTRKRYFH 375
 Db 301 LGMSQPCDVMWSIGCIIIEYVIGFTVFOTHDSKEHLAMMERILGPIPOHMIQKTRKRYFH 360
 QY 376 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKLPDLVRLMEYDPTORITLDEALQ 435
 Db 361 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKLPDLVRLMEYDPTORITLDEALQ 420
 QY 436 HPEFDLKK 444
 Db 421 HPEFDLKK 429

RESULT 5
 US-09-457-040B-12
 ; Sequence 12, Application US/09457040B
 ; Patent No. 6387641
 ; GENERAL INFORMATION:
 ; APPLICANT: Vertex Pharmaceuticals Incorporated
 ; APPLICANT: Bellon, Steve
 ; TITLE OF INVENTION: Crystallized P38 Complexes
 ; FILE REFERENCE: VPI/98-14

; CURRENT APPLICATION NUMBER: US/09/457,040B
 ; CURRENT FILING DATE: 1999-12-08
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-457-040B-12

Query Match 78.3%; Score 1887; DB 4; Length 484;
 Best Local Similarity 82.1%; Pred. No. 6,2e-169;
 Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY 18 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 77
 Db 54 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 113
 QY 78 KKR-RNRKSSH-QSRKSHRRKRSRSIIDEDEGHLICQSGDVLARAYEIVDTLGEAGF 135
 Db 114 KSKRIRHSHSTRSHSGSHRRKRTSRVEDDEGHLICQSGDVLARAYEIVDTLGEAGF 173
 QY 136 KVECIDHGMQDMHVAVKYVNGVYRPAARSEIOVLEHNLSTDPNSVFRVCQMLEFSDH 195
 Db 174 KVECIDHKGAGRHVAVKYVNGVYRPAARSEIOVLEHNLSTDPNSVFRVCQMLEFSDH 233
 QY 196 HGVCIIVFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPN 255
 Db 234 HGVCIIVFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPN 293
 QY 256 ILFVKSQDYVVKNSKMKRDERTLKNTDIAKVDPGSATYDDEHSTLVSTRHYRAPEVILA 315
 Db 294 ILFVKSQDYVVKNSKMKRDERTLKNTDIAKVDPGSATYDDEHSTLVSTRHYRAPEVILA 353
 QY 316 LGMSQPCDVMWSIGCIIIEYVIGFTVFOTHDSKEHLAMMERILGPIPOHMIQKTRKRYFH 375
 Db 354 LGMSQPCDVMWSIGCIIIEYVIGFTVFOTHDSKEHLAMMERILGPIPOHMIQKTRKRYFH 413
 QY 376 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKLPDLVRLMEYDPTORITLDEALQ 435
 Db 414 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKLPDLVRLMEYDPTORITLDEALQ 473
 QY 436 HPEFDLKK 444
 Db 474 HPEFDLKK 482

RESULT 6
 US-09-457-040B-13
 ; Sequence 13, Application US/09457040B
 ; Patent No. 6387641
 ; GENERAL INFORMATION:
 ; APPLICANT: Vertex Pharmaceuticals Incorporated
 ; APPLICANT: Bellon, Steve
 ; TITLE OF INVENTION: Crystallized P38 Complexes
 ; FILE REFERENCE: VPI/98-14
 ; CURRENT APPLICATION NUMBER: US/09/457,040B
 ; CURRENT FILING DATE: 1999-12-08
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 499
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-457-040B-13

Query Match 58.8%; Score 1417; DB 4; Length 499;
 Best Local Similarity 59.7%; Pred. No. 9,3e-125;
 Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;

QY 18 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 77
 Db 50 SHYLEARSLNERDYRRYVDEYRNDYCEGYVPRHYHRLDSGYRHCSSSVRSRSP 109

Qy	72	SRSSPKP-KRNRHCSQSSRSKRRRRNSIEDDEGHLCOSSGVNARAREIVDTL	129
Db	110	SQSSRRKRRRRRRRRRTFSRSSQSSRRASVEDDABGHLYHVGMLOERYEIVSTL	165
Qy	130	GEGAFCQVCEIDHGMDOGNHVAVKLYKNGVREARASEI OYLIELHNSIDNSVFCVOM	189
Db	170	GEGTFGVGVQCVDRHRRGGARVALKIIKNVEKKEARAREINVLKNEKNDPNKILCVOM	229
Qy	190	LEWFDHHGVCIVFELLGLSTYDFEIKENSFLRPOIDHIRQANAYQICOSINFLHNKULTHT	249
Db	230	FDWFDYGHMCISFELLGLSTDFPKONNYLTPYPIHQVHNAFOLCQVFKLHNKULTHT	289
Qy	250	DLMPENILFXSDSYVVKVSKMKRDEPRLKXNDICVNPFGSATYDDHEHSTLVSTRYRA	309
Db	290	DLRPEILFLPNSDYELTYLLEKKRDEBSYKSAFVAVVDFGSATFDHEHSTIVSTRYRA	349
Qy	310	PEVYIALGWSQPCDWSIGCILEYYLGGTVQOTDSKEHLAMMERILGPIPOQMIOKTR	369
Db	350	PEVILBELGMSQPCDWSIGCILEYYVGGFTLEQHNDNRHLLAMMERILGPIPSMIMIKTR	409
Qy	370	KRRYPHHNOLDWDEHSSAGRYVRRCKPLKEHMLCHDEHEKLPDLVARMLEYPDTORIT	429
Db	410	KQKYFRRGLDWDEHSSAGRYVRENCKPLRYLTSEAREHHQLDILJESMLEYPARLTL	469
Qy	430	LDLALQHPFRLDK 443	
Db	470	LGBALQHPFARLR 483	

FILE REFERENCE: GP-10218
CURRENT APPLICATION NUMBER: US/09/659,166
CURRENT FILING DATE: 2000-09-11
PRIOR APPLICATION NUMBER: UK 9921505.5
PRIOR FILING DATE: 1999-09-10
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT
ORGANISM: rattus
US-09-659-166-2

Query Match 22.8%; Score 550.5; DB 4; Length 527;
Best Local Similarity 34.3%; Pred. No. 2,5e-43;
Matches 121; Conservative 69; Mismatches 122; Indels 41; Gaps 7;

QY 105 DDEBGLICQSGDVLRAAYIVDTLGGAGFKVVECIDHGMGMHVAVKIKVNGRYREA 164
DB 130 DDDQGSYVQVPHDVAAYREVLKVIKGSFGQVYKAYDHKVH-QHVALKMYRNEKRFHRQ 188
QY 165 ARSEIQVLEHINSTDPNSVFRVOMLEWFDHGHVCIPELLGTYDFIKENSFLPFOI 224
DB 189 AAEIRILEHLRKQDKNTMNVIMLENFTRRNHICMTPELLSMNLVLIKKNKFGQFSI 248
QY 225 DHIRQAVYQICQINFLEHNNKLTHTDLKPENILFVKSQDYVVKYNSKMKRDERTLKNTDIK 284
DB 249 PLVRFKFAHSIIQCIDLALHKNRIIHCDLKPENIL-----LKQGR-----SGIK 291
QY 285 VDPGSAATYDDEHNSLTAVSTRHARPEVILALGMSQPCDWSIGCILIEYLGFTVPGTH 344
DB 292 VIDFSSCYEHQRYVYTIQSRFYRAPEVILGARVGMPIDMSLGCILAEILLTGPLPGE 351
QY 345 DSKEHLAMMERILGPIPOHMIQKTRKRYFHNNQ-----LDWDHSSAGRYVR 393
DB 352 DRGQDLACMIELGMPSQKLLDASKRAKNFVSXGPRYCTVTTISDGSVVLNGRSRSG 411
QY 394 RCKPLKE-----FMLCHDEHEKLF-DLYRMLEYDPTORTITDEALQHPF 438
DB 412 KLRGPRESREMGNAKKGCD-----PLFLDPLKQCLEWDPVAVMTFGQALRHPM 460

RESULT 10
US-08-802-466-2
Sequence 2, Application US/08802466
Patent No. 5972606
GENERAL INFORMATION:
APPLICANT: Creasy, et al.
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,466
FILING DATE: 19 February 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: William T. Han
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GH50002

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-802-466-2

Query Match 22.7%; Score 546.5; DB 2; Length 528;
Best Local Similarity 34.3%; Pred. No. 6e-43;
Matches 121; Conservative 68; Mismatches 123; Indels 41; Gaps 7;

QY 105 DDEBGLICQSGDVLRAAYIVDTLGGAGFKVVECIDHGMGMHVAVKIKVNGRYREA 164
DB 131 DDDQGSYVQVPHDVAAYREVLKVIKGSFGQVYKAYDHKVH-QHVALKMYRNEKRFHRQ 189
QY 165 ARSEIQVLEHINSTDPNSVFRVOMLEWFDHGHVCIPELLGTYDFIKENSFLPFOI 224
DB 190 AAEIRILEHLRKQDKNTMNVIMLENFTRRNHICMTPELLSMNLVLIKKNKFGQFSI 249
QY 225 DHIRQAVYQICQINFLEHNNKLTHTDLKPENILFVKSQDYVVKYNSKMKRDERTLKNTDIK 284
DB 250 PLVRFKFAHSIIQCIDLALHKNRIIHCDLKPENIL-----LKQGR-----SGIK 292
QY 285 VDPGSAATYDDEHNSLTAVSTRHARPEVILALGMSQPCDWSIGCILIEYLGFTVPGTH 344
DB 293 VIDFSSCYEHQRYVYTIQSRFYRAPEVILGARVGMPIDMSLGCILAEILLTGPLPGE 352
QY 345 DSKEHLAMMERILGPIPOHMIQKTRKRYFHNNQ-----LDWDHSSAGRYVR 393
DB 353 DRGQDLACMIELGMPSQKLLDASKRAKNFVSXGPRYCTVTTISDGSVVLNGRSRSG 412
QY 394 RCKPLKE-----FMLCHDEHEKLF-DLYRMLEYDPTORTITDEALQHPF 438
DB 413 KLRGPRESREMGNAKKGCD-----PLFLDPLKQCLEWDPVAVMTFGQALRHPM 461

RESULT 11
US-09-350-484-2
Sequence 2, Application US/09350484
Patent No. 6159716
GENERAL INFORMATION:
APPLICANT: Creasy, et al.
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,484
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/802,466
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: William T. Han
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GH50002


```

Qy 285 VYDFGSATYDDDEHSTLVSTRHYRAPEVILALGMSQPCDVWSIGCIIIEYVIGFTVFQTH 344
Db 353 VIDFGSSCFEYOKLYTYIQSRFYRAPEIILGSRYSTPIDIWSFGCIIAELITGQPLFPGH 412
Qy 345 DSKHEILAMMERITIGPIPOHMIQTRKRYFHNN-----QIDMDEHSSAGRYV-- 391
Db 413 DEGDQIASMMELGMPPKLBOSSKRAKYPINSKGIPRYCSVTTOAD-----GRVVLV 465
Qy 392 -----RRCKP-----LKEFMLCHDEHEKLF-DVVRMLEYDPTORITLDEALQ 435
Db 466 GGRSRRGKKRGPPGSKDWGTALKG--CD-----YLFIEFLKRCUHWBPARLITPAQALR 518
Qy 436 HPF 438
Db 519 HPW 521

```

Search completed: November 14, 2003, 15:19:48
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:15:24 ; Search time 21 Seconds
(without alignments)

2037,860 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410

Sequence: 1 MCIPLFASHSVSEEDTPSHY.....QRTLDALQHPFDLLKKK 445

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76.*
2: PIR1.*
3: PIR2.*
4: PIR3.*
5: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	78.3	484	2	protein kinase CLK
2	1837	76.2	483	2	protein kinase STY
3	1727	71.7	484	2	protein kinase (BC
4	1417	58.8	499	2	protein kinase CLK
5	1321	54.8	490	2	protein kinase CLK
6	1317	54.6	490	2	protein kinase CLK
7	1222	50.7	517	2	protein kinase CLK
8	965.5	40.1	903	2	protein kinase CLK
9	876	36.3	306	2	protein kinase CLK
10	848.5	35.2	427	2	protein kinase CLK
11	839	34.8	431	2	protein kinase CLK
12	830	34.4	467	2	protein kinase CLK
13	790	32.8	400	2	protein kinase CLK
14	767	31.8	575	2	protein kinase CLK
15	767	31.8	690	2	protein kinase CLK
16	718.5	29.8	737	1	protein kinase CLK
17	530	22.0	543	2	protein kinase CLK
18	524	21.7	1087	2	protein kinase CLK
19	523	21.7	508	2	protein kinase CLK
20	523	21.7	817	2	protein kinase CLK
21	504.5	20.9	589	2	protein kinase CLK
22	478.5	19.9	629	2	protein kinase CLK
23	477	19.8	1157	2	protein kinase CLK
24	474	19.7	948	2	protein kinase CLK
25	472.5	19.6	754	2	protein kinase CLK
26	448.5	18.6	1189	2	protein kinase CLK
27	443.5	18.4	1155	2	protein kinase CLK
28	441	18.3	1192	2	protein kinase CLK
29	437	18.1	642	2	protein kinase CLK

30	435	18.0	570	2	probable protein k
31	432	17.9	1191	2	serine/threonine p
32	431.5	17.9	807	2	protein kinase YAK
33	430	17.8	821	2	hypothetical prote
34	429	17.8	1457	2	protein kinase YAK
35	419	17.4	1209	2	homeodomain-intera
36	406.5	16.9	477	2	mRNA splicing-asso
37	401	16.6	1093	2	protein B0464.5a l
38	398	16.5	775	2	hypothetical prote
39	392	16.3	1087	2	hypothetical prote
40	389	16.1	440	2	probable protein k
41	380.5	15.8	438	2	protein kinase hom
42	367.5	15.2	782	2	protein B0464.5c l
43	343	14.2	333	2	protein kinase (EC
44	338.5	14.0	523	2	serine protein kin
45	334	13.9	361	2	mitogen-activated

ALIGNMENTS

RESULT 1

S53641
protein kinase CLK (EC 2.7.1.1) - human

C/Species: Homo sapiens (man)

C/Date: 15-Jul-1995 #sequence #revision 01-Sep-1995 #text_change 03-May-2002

C/Accession: S53641

R/Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.

J. Mol. Biol. 244, 665-672, 1994

A/Title: Characterization by CDNA cloning of two new human protein kinases. Evidence b

A/Reference number: S53637; M01D:95082033; PMID:7990150

A/Accession: S53641

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Cross-references: GB:U29219; NID:9632963; PTD:AAA61480.1; PTD:9632964

C/Superfamily: human protein kinase CLK1; protein kinase homology

C/Keywords: alternative splicing; phosphotransferase

F:159-430/Domain: protein kinase homology <Kin>

Query Match 78.3%; Score 1887; DB 2; Length 484;

Best Local Similarity 82.1%; Pred. No. 5e-83;

Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY	18	SHYLAASLNERNDRYVDEYRNDYCEGYPRHYHNDISGRYHSSKSSVGRSSP	77
DB	54	SHYLSRSINEXDYHRRYIDEYRNDYQGEFGHQRQDHESRYQNHSSKSSGRSSY	113
QY	78	KKK-RNRHCSH-QSRSKSRKRSRSTEDDEBGLICQSGDVLARYEIVDTLGEAFG	135
DB	114	KSKRHINHSRSHGSHRKRKTRSVEDDEBGLICQSGDVLARYEIVDTLGEAFG	173
QY	136	KVVECIDHMDGMHVAVKIVKVGGRYRPARSEIOVLEHLSNSTDPSVFRVQMLEFDD	195
DB	174	KVVECIDHMDGMHVAVKIVKVGGRYRPARSEIOVLEHLSNSTDPSVFRVQMLEFDD	233
QY	196	HEHVCIVFELGLSYDFIKNSPLPOIDHIRQAAVQICQINFLHNKLTHTDLKEN	255
DB	234	HEHVCIVFELGLSYDFIKNSPLPOIDHIRQAAVQICQINFLHNKLTHTDLKEN	293
QY	256	ILFVSDYVVKYNSPKKDEKTLKNTDIKVDYFGSATDDEHSTLVSTRYRAPEVILA	315
DB	294	ILFVSDYVVKYNSPKKDEKTLKNTDIKVDYFGSATDDEHSTLVSTRYRAPEVILA	353
QY	316	LGMSQPCVWSIGCLILEYVGFVYFQHDSEKHAMMERILGPIPGMIQTRKKFFH	375
DB	354	LGMSQPCVWSIGCLILEYVGFVYFQHDSEKHAMMERILGPIPGMIQTRKKFFH	413
QY	376	HNOLDMDHSSAGRYVRRCKPLKEFMLCHDEBEKLPDLVRLMEIYPTORTIDEALQ	435
DB	414	HNOLDMDHSSAGRYVRRCKPLKEFMLCHDEBEKLPDLVRLMEIYPTORTIDEALQ	473
QY	436	HEFDLLKK 444	

Db 474 HPFDLLKK 482

RESULT 2

A39676 protein kinase STY (EC 2.7.1.-) - mouse

N/Alternate names: protein kinase clk

C/Species: Mus musculus (house mouse)

C/Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 03-May-2002

C/Accession: A39676; S13364

R/Howell, B.W.; Afar, D.E.H.; Lew, J.; Douville, E.M.J.; Icelly, P.L.E.; Gray, D.A.; Bell, M.J. Cell Biol. 11, 568-572, 1991

A/Title: STY, a tyrosine-phosphorylating enzyme with sequence homology to serine/threonine

A/Reference number: A39676; PMID:1986248

A/Accession: A39676

A/Molecule type: mRNA

A/Residues: 1-483 <HOM>

A/Cross-references: GB:M38381, NID:g201070; PIDN:AAA0151.1; PID:g201071

R/Ben-David, Y.; Letwin, K.; Tamock, L.; Bernstein, A.; Pawson, T.

EMBO J. 10, 317-325, 1991

A/Title: A mammalian protein kinase with potential for serine/threonine and tyrosine pho

A/Reference number: S13364; PMID:91122038; PMID:1825055

A/Accession: S13364

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-378, 'P', 380-483 <BEN>

A/Note: the sequence from Fig. 2 is inconsistent with that shown in Fig. 1 in having 448

C/Superfamily: human protein kinase clk; protein kinase homology

C/Keywords: phosphotransferase; serine/threonine-specific protein kinase

F:158-429/Domain: protein kinase homology <KIN>

Query Match 76.2%; Score 1837; DB 2; Length 483;

Best Local Similarity 79.7%; Pred. No. 1.2e-80;

Matches 349; Conservative 32; Mismatches 51; Indels 6; Gaps 4;

```

QY 9 HSEVEDTHSHYLSERDSDRYVDEYNDYCEGYVPRHYRDIESGRIHCSGS 68
DB 48 HSKTID--SYLSESSINIEKAYHSRRYVDEYNDYM-GYEGHPGEGSRYQWMSKXS 103
QY 69 SVSRSSSPKRR-RNR-HCSHQSSRSKSHRRKRSRIEDDEEGHLICQSGDVLARYEIV 126
DB 104 SGRSGSSYSKSGHRRHHTSQHSHQSKSHRRKRSVEDDEEGHLICQSGDVLARYEIV 163
QY 127 DTLSGARGKVECTIDHGMVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRC 186
DB 164 DTLSGARGKVECTIDHGMVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRC 223
QY 187 VQMLMPFHGHVCIIVFELIGSTYDFIKENSFLPQIDHIRQMAVOICQINFLHNNKL 246
DB 224 VQMLMPFHGHVCIIVFELIGSTYDFIKENSFLPQIDHIRQMAVOICQINFLHNNKL 283
QY 247 THTDLPENILFVKSDDYVVKYNSKMKRDETLKNTDIAKVDFGSATYDDEHSTVSTRH 306
DB 284 THTDLPENILFVKSDDYVVKYNSKMKRDETLKNTDIAKVDFGSATYDDEHSTVSTRH 343
QY 307 YRAPVILALGMSQCDVMSIGCIIIEYLGLFTVQTHDSKHEMLAMERILGPIPOHMQ 366
DB 344 YRAPVILALGMSQCDVMSIGCIIIEYLGLFTVQTHDSKHEMLAMERILGPIPOHMQ 403
QY 367 KTRRRKYFHHNQLDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFPLVRRMLEYDPTQ 426
DB 404 KTRRRKYFHHNQLDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFPLVRRMLEYDPTQ 463
QY 427 RITLDEALQHPFDLLKK 444
DB 464 RITLDEALQHPFDLLKK 481

```

RESULT 3

A38643

protein kinase (EC 2.7.1.37) cdc2/cdc28-like - human

C/Species: Homo sapiens (man)

C/Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 03-May-2002

C/Accession: A38643

R/Johnson, K.W.; Smith, K.A.

J. Biol. Chem. 266, 3402-3407, 1991

A/Title: Molecular cloning of a novel human cdc2/CDC28-like protein kinase.

A/Reference number: A38643; PMID:1704889

A/Accession: A38643

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-454 <JOH>

A/Cross-references: GB:M59287

C/Genetics:

A/Map position: 7q31-7q31

C/Superfamily: human protein kinase clk; protein kinase homology

C/Keywords: phosphotransferase

F:129-400/Domain: protein kinase homology <KIN>

Query Match 71.7%; Score 1727; DB 2; Length 454;

Best Local Similarity 76.6%; Pred. No. 1.9e-75;

Matches 327; Conservative 26; Mismatches 46; Indels 28; Gaps 3;

```

QY 18 SHYLEARSINERDSDRYVDEYNDYCEGYVPRHYRDIESGRIHCSGSVSRSSP 77
DB 54 SHYLEARSINERDSDRYVDEYNDYCEGYVPRHYRDIESGRIHCSGSVSRSSP 113
QY 78 KKKRNRHCSHQSSRSKSHRRKRSRIEDDEEGHLICQSGDVLARYEIVDTLGGAFGKV 137
DB 114 KKKRNRHCSHQSSRSKSHRRKRSRIEDDEEGHLICQSGDVLARYEIVDTLGGAFGKV 145
QY 138 VECIDHGMVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRVQMLEWHDHG 197
DB 146 VECIDHGMVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRVQMLEWHDHG 205
QY 198 HVCIVFELIGSTYDFIKENSFLPQIDHIRQMAVOICQINFLHNNKLTHDKEENIL 257
DB 206 HVCIVFELIGSTYDFIKENSFLPQIDHIRQMAVOICQINFLHNNKLTHDKEENIL 265
QY 258 FYKSDYVVKYNSKMKRDETLKNTDIAKVDFGSATYDDEHSTVSTRHRAPEVILALG 317
DB 266 FYKSDYVVKYNSKMKRDETLKNTDIAKVDFGSATYDDEHSTVSTRHRAPEVILALG 325
QY 318 WSQPCDVMSIGCIIIEYLGLFTVQTHDSKHEMLAMERILGPIPOHMQTKTKRYFHHN 377
DB 326 WSQPCDVMSIGCIIIEYLGLFTVQTHDSKHEMLAMERILGPIPOHMQTKTKRYFHHN 385
QY 378 QLDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFPLVRRMLEYDPTQITLDEALQHP 437
DB 386 QLDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFPLVRRMLEYDPTQITLDEALQHP 445
QY 438 FFDLLKK 444
DB 446 FFDLLKK 452

```

RESULT 4

S53637

protein kinase clk2, long splice form (EC 2.7.1.-) - human

C/Species: Homo sapiens (man)

C/Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 03-May-2002

C/Accession: S53637; T08825

R/Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.

J. Mol. Biol. 244, 665-672, 1994

A/Title: Characterization by cDNA cloning of two new human protein kinases. Evidence by

A/Reference number: S53637; PMID:95082033; PMID:7990150

A/Accession: S53637

A/Molecule type: mRNA

A/Residues: 1-499 <HAN>

A/Cross-references: GB:I29218; NID:9632967; PIDN:AAA61482.1; PID:9632968

R/Minfeld, S.L.; Tayebi, N.; Martin, B.M.; Gimms, E.I.; Sidransky, B.

Genome Res. 7, 1020-1026, 1997

A/Title: Identification of three additional genes contiguous to the glucocorticoid receptor

A:Reference number: 216482; MUID:97474796; PMID:9331372
 A:Accession: T08825
 A>Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-499 <MIN>
 A:Cross-references: GB:AF023268; NID:92564910; PIDN:AA51817.1; PID:92564911
 C:Comment: The short splice form of this protein (see PIR:S53638) lacks the protein kinase domain.
 C:Gene: CLK2
 C:Function: Phosphorylates Ser, Thr, and Tyr residues on proteins in the spliceosomal C/superfamily: human protein kinase CLK1; protein kinase homology
 C:Keywords: alternative splicing; phosphotransferase
 F:161-440/Domain: protein kinase homology <KIN>

Query Match 58.8%; Score 1417; DB 2; Length 499;
 Best Local Similarity 59.7%; Pred. No. 1; Le-60;
 Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;

```

18 SHYLEARS-----LNRDVRDRRYVDEY-RNDYCEGVPRHHRDIESGTRHCSKSVR 71
Db 50 SHVRSRSSYDDSSDRRYDRRYCGSYRNDYSRDGDAYDTDYRHSTYEQRENSSYR 109
QY 72 SRRSSPK-RKENRHCSSHQSRKSHRRKRSIEDDEEGHLICQSGDYLARAYIVDTL 129
Db 110 SRRSSPKRRRRRRRRRTSRSSQHSRRKSVEDAECHLIYHGDMLQERYELIVSTL 169
QY 130 GSGAEKRVETCDHGMGNHVAIKYKNGRYEARSETQVLEHLNSTDPNSVFCVOM 189
Db 170 GSGTFRGVQCVDRHREGARVALKIKYNEKYKEARLEINVLKINEDKPDKNLCYQM 229
QY 190 LEMFDHGHVCIWFELIGLSTYDEIKENSFLPRQIDITQMAVQIQSINFLHANKLTHT 249
Db 230 FEMFDHGHMCISFELIGLSTFELDKNNYLPRPHQVHMAFQLCAVAKFHDNCLTHT 289
QY 250 DLKPNILFVKSDYVVKYNSKMKRDERTLKNTDIXKVDGSAFYDEHSTLVSTRHYA 309
Db 290 DLKPNILFVNSDYELTYLTKKRDERSYKSTAVRVYDGSATFDEHSTLVSTRHYA 349
QY 310 PEVILALGWSQPCDWSIGICILLEYLIGFTVQTHDSKEHLAMMERILGPIPDHMIQTR 369
Db 350 PEVILALGWSQPCDWSIGICILFEYVYGFTLQTHDNREHLAMMERILGPIPSRMIRKTR 409
QY 370 KRKYFHNLQDWDHSSAGRYRRCKPKKEPMLCHDEHEKLFDLVRMLEYDPTQRI 429
Db 410 KRKYFGRGLDWDHSSAGRYRCKPLRLRYLTSBAEHLQDLFDEISMLEYEPKRLT 469
QY 430 IDEALQHPFDLLK 443
Db 470 LGBALQHPFFARLR 483

```

RESULT 5
 S53639
 protein kinase CLK3 (EC 2.7.1.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 03-May-2002
 C:Accession: S53639; S71040
 R:Haes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.
 J. Mol. Biol. 244, 665-672, 1994
 A:Title: Characterization by cDNA cloning of two new human protein kinases. Evidence by
 A:Reference number: S53637; MUID:95082033; PMID:7990150
 A:Accession: S53639
 A:Molecule type: mRNA
 A:Residues: 1-490 <HAN>
 A:Cross-references: GB:L29217
 R:Haes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.
 Submitted to the EMBL Data Library, January 1995
 A:Reference number: S71040
 A:Accession: S71040
 A:Molecule type: mRNA
 A:Residues: 1-131; 'TG', 134-490 <HAN>
 A:Cross-references: EMBL:L29217; NID:9632971; PIDN:AAA61484.1; PID:9632972

C:Superfamily: human protein kinase CLK1; protein kinase homology
 C:Keywords: alternative splicing; phosphotransferase
 F:154-433/Domain: protein kinase homology <KIN>

Query Match 54.8%; Score 1321; DB 2; Length 490;
 Best Local Similarity 55.3%; Pred. No. 3; 7e-56;
 Matches 251; Conservative 60; Mismatches 101; Indels 42; Gaps 6;

```

4 PLEASHVEEDTPSHYLEARSINERDYDRRYVDEYRNDYCEGYVP-----R 51
Db 43 PPRRSRSRSHDRP-----YQRRYRRRSDTYR---CEERSFPEDYGGPSSR 90
QY 52 HYHRIEBC-YRI-----HCKSSVRSRRSPKRRNHCSSHQSRKSHRRKRSRSTED 105
Db 91 HRRRSRERGPYRTRKHAHCHK-----RTRSCSSASRSRQSSKSSRSRVED 138
QY 106 DEEGHLICQSGDYLARAYIVDTLGEAGKRVETCDHGMGNHVAIKYKNGRYEAR 165
Db 139 DKEGHLVCRIGDMLQERYELIVGNLGGTGKVVCCIDHARGKQVALKIRNGKYREAA 198
QY 166 RSEIQLVLEHLNSTDPNSVFCVOMLEMPDHGHVCIWFELIGLSTYDEIKENSFLPRQID 225
Db 199 RLEINVLKIKKEDKXNKFLCVLSMDYFPHGMCLAFELIGKNTFEFLKNNPQYPLP 258
QY 226 HIRQAAVQIQSINFLHANKLTHTDIKPENILFVKSDYVVKYNSKMKRDERTLKNTDIX 285
Db 259 HVRHMAVQICHALFLENQLTHTDLKPNILFVNSEFETLVNHSKCEKSYVNTSIRV 318
QY 286 VFGSATYDDEHSTLVSTRHARPEVILALGWSQPCDWSIGICILLEYLIGFTVQTHD 345
Db 319 ADGSAFVPEHHTTIVATRRHYRPEVILLELGAQPCDWSIGICILFEYRGGFTLFTQTH 378
QY 346 SKHELTAMMERILGPIPDHMIQTRKRYFHHNOLDHSSAGRYRRCKPKKEPMLCH 405
Db 379 NREHLVMEKILGPIPSHILHTRKQKYYKGGVLWDHSSDGRYVKNCKPLKSYMLQD 438
QY 406 DEEHEKLFDLVRMLEYDPTQRIIDEALQHPFF 439
Db 439 SLEHVQLFDLMRMLFEDPAQRITLAEALLHPFF 472

```

RESULT 6
 S70352
 protein kinase CLK3 (EC 2.7.1.-), testis-specific - rat
 N:Alternate names: LAMMER kinase
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 03-May-2002
 C:Accession: S70352
 R:Becker, W.; Kentrup, H.; Heukelbach, J.; Joost, H.G.
 Biochim. Biophys. Acta 1312, 63-67, 1996
 A:Title: cDNA cloning and characterization of rat CLK3, a LAMMER kinase predominately
 A:Reference number: S70352; MUID:96271481; PMID:8679717
 A:Accession: S70352
 A:Molecule type: mRNA
 A:Residues: 1-490 <BEC>
 A:Cross-references: EMBL:X94351; NID:91149536; PIDN:CAA64076.1; PID:91149537
 C:Gene: CLK3
 C:Superfamily: human protein kinase CLK1; protein kinase homology
 C:Keywords: phosphotransferase; protein kinase
 F:154-433/Domain: protein kinase homology <KIN>

Query Match 54.6%; Score 1317; DB 2; Length 490;
 Best Local Similarity 55.1%; Pred. No. 5; 8e-56;
 Matches 250; Conservative 61; Mismatches 101; Indels 42; Gaps 6;

```

4 PLEASHVEEDTPSHYLEARSINERDYDRRYVDEYRNDYCEGYVP-----R 51
Db 43 PPRRSRSRSHDRP-----YQRRYRHRSDTYR---CEERSFPGEPCYSSSR 90
QY 52 HYHRIEBC-YRI-----HCKSSVRSRRSPKRRNHCSSHQSRKSHRRKRSRSTED 105
Db 91 HRRRSRERGPYRTRKHAHCHK-----RTRSCSSASRSRQSSKSSRSRVED 138

```


Db 405 RFPDSIRMTAHDAALRHPHF 423

RESULT 12

S71169

protein kinase, 54K (EC 2.7.1.-) - Arabidopsis thaliana

N:Alternate names: protein F4P12.270

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 03-May-2002

C/Accession: S71169, T45897

R:Kurumori, T.; Yamamoto, M.

submitted to the EMBL Data Library, January 1995

A:Description: A.thaliana genes encoding protein kinases of a new family.

A:Reference number: S71169

A:Accession: S71169

A:Molecule type: mRNA

A:Residues: 1-467 <KUR>

A:Cross-references: EMBL:D45354; NID:g642131; PID:g642132

R:Blocker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23016

A:Accession: T45897

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-467 <BLO>

A:Cross-references: EMBL:AL132966

A:Experimental source: cultivar Columbia; BAC clone F4P12

C:Genetics:

A:Map position: 3

A:Introns: 56/3; 115/1; 124/1; 175/2; 191/3; 230/1; 260/3; 313/1; 334/3; 367/2; 402/3

C:Superfamily: human protein kinase c1k1; protein kinase homology

C:Keywords: ATP; phosphotransferase; protein kinase

F:113-401/Domain: protein kinase homology <KIN>

Query Match 34.4%; Score 830; DB 2; Length 467;

Best Local Similarity 45.8%; Pred. No. 7.2e-33;

Matches 160; Conservative 71; Mismatches 102; Indels 16; Gaps 8;

105 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCECIDHGMGMVAIVKIVNGVREAA 164

Db 97 DDKDGHVYFVGGDTLPRYQILSKYGEFGVLCBEPDNKNEV-VAIIVISINKYREA 155

QY 165 ARSEIQVLEHLNSTDPNSVFCVOMLEWFDHGHVCIYFELLGLSTYDFIKNSFLPQI 224

Db 156 AMIEIDVILQRLTRHVG-SRCVQIRNMPDYRNHICIVFEKLGPSLDFLRKNSRSPFI 214

QY 225 DHROMAYOICQSIINFLHNKLTHTDLKPENLIFVKSQDV---VKYNSKMKRDEFTLKN 280

Db 215 DLVREIGRLQLESVAVMMDLRILHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 274

QY 281 ---TDIKVDFGSATYDDEHSTLVSTRHYAPFVITLALGMSOPCDVWSIGCILEYYL 336

Db 275 LFKSAIKLIDFGSTTFEHOQNIYVSTRHYAPFVITLALGMSOPCDVWSIGCILEVELS 334

QY 337 GTTVQTHDSKSHLAMEKILPPIQPHMIQKT-RKRYFHHN-QLDWDHSSAGRYVR 393

Db 335 GEALQTHNEHNLAMERVLGLPPLPHVTLRADRSSEKFRGAKLDWEGATSRDSIXA 394

QY 394 --RCKPLKEFMLCH-DEEHEKLFDLVRMLQYDPTQRIITLDAALQHPF 439

Db 395 VWKLPRLNLMQHDHSDAGLLDILQGLRKYDPTFERKAREALNHPF 443

RESULT 13

T04460

protein kinase AMB3 (EC 2.7.1.-) - Arabidopsis thaliana

N:Alternate names: protein F4D11.140

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Mar-2000

C/Accession: T04460

R:Bevan, M.; Benes, V.; Rechmann, S.; Borikova, D.; Ansoorge, W.; Hohelsel, J.; Mewes, H.W.

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15360

A:Accession: T04460

A:Molecule type: DNA

A:Residues: 1-400 <BEV>

A:Cross-references: EMBL:AL022537

A:Experimental source: cultivar Columbia; BAC clone F4D11

C:Genetics:

A:Gene: AMB3

A:Map position: 4

A:Introns: 30/3; 71/1; 80/1; 131/2; 147/3; 186/1; 216/3; 266/1; 287/3; 320/2; 355/3

A:Note: F4D11.140

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homo

C:Keywords: ATP; phosphotransferase; protein kinase

Query Match 32.8%; Score 790; DB 2; Length 400;

Best Local Similarity 41.7%; Pred. No. 4.9e-31;

Matches 169; Conservative 70; Mismatches 134; Indels 32; Gaps 11;

QY 57 ISSGYRHCSKSVSR-----RSSPKRR--NRCSHQSKSHRRKRSIED 105

Db 2 IANGFE-SMDKRRVKKRPMTDDEAPAEPAKRAVYKGGSDGRILSPPLR-----DD 53

QY 106 DEEGHLICQSGDVLARAYEIVDTLGEAGFVCECIDHGMGMVAIVKIVNGVREAA 165

Db 54 DRDGHVFEIIRDLNLTFRYKILSKMGEGFVLECDWDRDK-EYVAIKIIRSIKKYRDA 112

QY 166 RSEIQVLEHLNSTDPNSVFCVOMLEWFDHGHVCIYFELLGLSTYDFIKNSFLPQI 225

Db 113 MEIDVILQRLTRHVG-SRCVQIRNMPDYRNHICIVFEKLGPSLDFLRKNSRSPFI 171

QY 226 DHROMAYOICQSIINFLHNKLTHTDLKPENLIFVKSQDVYVKNKSKKROBERT---LKN 280

Db 172 LVDRDGCQLLEBVAIMEIQLVHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 231

QY 281 TDIKVDFGSATYDDEHSTLVSTRHYAPFVITLALGMSOPCDVWSIGCILEYYLFTV 340

Db 222 SAIKLIDFGSTTCNDRHNSIVQTRHYSPEVITLALGMSOPCDVWSIGCILEVELS 291

QY 341 POTHDSKSHLAMEKILPPIQPHMIQKT-RKRYFHHN-QLDWDHSSAGRYVR-RRC 395

Db 232 FQTHNLHLAMERVLGLPPLPHVTLRADRSSEKFRGAKLDWEGATSRDSIXA 351

QY 396 RCKPLKEFMLCH-DEEHEKLFDLVRMLQYDPTQRIITLDAALQHPF 439

Db 352 DRKQWVSKHVDNTRSPADLLYGLIADPSERLTANRALDHPF 396

RESULT 14

JC7794

lammer kinase homolog protein 1, Lkh1 protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 02-Apr-2002

C/Accession: JC7794

R:Kim, K.H.; Cho, Y.M.; Kang, W.H.; Kim, J.H.; Byun, K.H.; Park, Y.D.; Bae, K.S.; Park,

Biochem. Biophys. Res. Commun. 289, 1237-1242, 2001

A:Title: Negative regulation of filamentous growth and flocculation by Lkh1, a fission

A:Reference number: JC7794

A:Accession: JC7794

A:Molecule type: mRNA

A:Residues: 1-575 <KIM>

A:Cross-references: GB:AF334941

A:Experimental source: strain SP286, ED665

C:Comment: This protein, a homolog of the dual-specificity protein kinase of the LAMMER

ast, by repressing expression of cell surface proteins for flocculation and morphogenesis

C:Genetics:

A:Gene: Lkh1+

Query Match 31.8%; Score 767; DB 2; Length 575;

Best Local Similarity 41.5%; Pred. No. 8.5e-30;

Matches 141; Conservative 78; Mismatches 115; Indels 6; Gaps 3;

QY 105 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCECIDHGMGMVAIVKIVNGVREAA 164

Db 113 MEIDVILQRLTRHVG-SRCVQIRNMPDYRNHICIVFEKLGPSLDFLRKNSRSPFI 171

[illegible]

probable protein kinase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C:Accession: T38052
R:Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, February 1995
A:Reference number: Z21765
A:Accession: T38052
A:Stratus: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1,690 <1kb>
A:Cross-references: EMBL:Z69239, PIDN:CA93220.1; GSPDB:GM00066, SPDB:SPAC1D4.11d
A:Experimental source: strain 972h-; cosmid CID4
C:Genetics:
A:Gene: SPDB:SPAC1D4.11c
A:Map position: 1
C:Superfamily: yeast probable protein kinase KNS1; protein kinase homology

Query Match	31.8%;	Score 767;	DB 2;	Length 690;
Best Local Similarity	41.5%;	Pred. No. 1e-29;		
Matches 141;	Conservative 78;	Mismatches 115;	Indels 6;	Gaps 3;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:09:04 ; Search time 17 seconds

(without alignments)
1230.993 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410

Sequence: 1 MCIPLEASHSVEDTHPSHY.....QRTLDALQHPFDLLKKK 445

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2261.5	93.8	481	1	CLK4 MOUSE
2	1887	78.3	484	1	CLK1 HUMAN
3	1832	76.0	483	1	CLK1 MOUSE
4	1427.5	59.2	499	1	CLK2 MOUSE
5	1417	58.8	499	1	CLK2 HUMAN
6	1321	54.8	490	1	CLK3 HUMAN
7	1317	54.6	490	1	CLK3 RAT
8	1310	54.4	490	1	CLK3 MOUSE
9	1222	50.7	517	1	DOA DROME
10	848.5	35.2	427	1	APC2 ARATH
11	830	34.4	467	1	APC1 ARATH
12	790	32.8	400	1	APC3 ARATH
13	767	31.8	575	1	LKHL SCHEP
14	718.5	29.8	737	1	KNS1 YEAST
15	549.5	22.8	528	1	DYR2 HUMAN
16	536.5	22.3	588	1	DYR3 HUMAN
17	530	22.0	836	1	KAB1 SCHPO
18	524	21.7	1087	1	POM1 SCHPO
19	517	21.5	541	1	DYR4 HUMAN
20	505	21.0	843	1	KMB DROME
21	500.5	20.8	589	1	DYR5 HUMAN
22	495.5	20.6	624	1	DYR3 DROME
23	478.5	19.9	629	1	DYR6 HUMAN
24	472.5	19.6	763	1	DYR4 HUMAN
25	472.5	19.6	763	1	DYR4 MOUSE
26	472.5	19.6	763	1	DYR4 RAT
27	451.5	18.7	1198	1	HLK2 HUMAN
28	450.5	18.7	1196	1	HLK2 MOUSE
29	431.5	17.9	807	1	YAK1 YEAST
30	422.5	17.5	1007	1	PR4B MOUSE
31	420	17.4	1007	1	PR4B HUMAN
32	406.5	16.9	477	1	PRP4 SCHPO
33	401	16.6	1093	1	YKD5_CABEL

34	359	14.9	533	1	ST23 HUMAN	Q9261 homo sapien
35	343	14.2	353	1	FUS3 YEAST	P16892 saccharomyc
36	335.5	13.9	348	1	M14B BRARE	Q9601 brachydanio
37	335.5	13.9	1030	1	STK9 HUMAN	O76039 homo sapien
38	334	13.9	565	1	ST23 MOUSE	Q92092 mus musculu
39	332.5	13.8	361	1	M14A CYPCA	Q90336 cypripinus ca
40	328.5	13.6	361	1	M14B CYPCA	Q91858 cypripinus ca
41	326.5	13.5	360	1	MK14 HUMAN	Q16539 h mitogen-a
42	326.5	13.5	360	1	MK14 PANTR	Q95927 pan troglod
43	325.5	13.5	360	1	MK14 CANFA	O02812 canis famill
44	323.5	13.4	360	1	MK14 MOUSE	P47811 mus musculu
45	322.5	13.4	360	1	MK14 RAT	P70618 rattus norv

ALIGNMENTS

RESULT 1
ID CLK4_MOUSE STANDARD; PRT; 481 AA.
AC C35493;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein kinase CLK4 (EC 2.7.1.1).
GN CLK4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97439710; PubMed=9307018;
RA Naylor O., Stamm S., Ullrich A.;
RT "Characterization and comparison of four serine- and arginine-rich
RT (SR) protein kinases."
RL Biochem. J. 326:693-700(1997).
CC -!- FUNCTION: PHOSPHORYLATES SERINE- AND ARGinine-RICH (SR) PROTEINS
CC OF THE SPICEOSOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- PTM: Autophosphorylates on all three types of residue.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC LAMMER SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF035566; AAB87510.1; -
CC HSSP; 000534; 1B18.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SMO0220; S_TKc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW transferase; Serine/threonine-protein kinase; ATP-binding;
KW tyrosine-protein kinase; Phosphorylation; Nuclear protein.
FT DOMAIN 159 475
FT NP_BIND 165 173
FT BINDING 189 189
FT ACT_SITE 286 286
FT ACT_SITE 286 286
SQ SEQUENCE 481 AA; 57344 MW; F2C56965900C12AA CRC64;


```

Query Match      93.8%; Score 2261.5; DB 1; Length 481;
Best Local Similarity 93.8%; Pred. No. 5,7e-137;
Matches 420; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

QY 8 SH-SVEEDTH-----PSHYLARSILNERDYRDRRYVDEYRNDYCEGYVRHHRD 56
D 33 SHSSTQENRCHCKPHHQKDSCHLHLKRCILNERDYRDRRYVDEYRNDYCEGYVRHHRD 92
QY 57 IESGYRHHCKSSYRSHRSRSPKRRKRNPHCSHQSRSKSHRRKRSIJEDEBGLICQSG 116
D 93 VESTYRHHCKSSYRSHRSRSPKRRKRNPHCSHQSRSKSHRRKRSIJEDEBGLICQSG 152
QY 117 DVLRARYEIVDTLGEAGFGKVECIDHGMGMHVAVKIVKNGYRRAASEIOVLEHLN 176
D 153 DVLRARYEIVDTLGEAGFGKVECIDHGMGMHVAVKIVKNGYRRAASEIOVLEHLN 212
QY 177 STDPSNYSRCVOMLEWFDHGHVCIPELLGLSTYDPIKENSFLPFOIDHROMAYOICQ 236
D 213 STDPSNYSRCVOMLEWFDHGHVCIPELLGLSTYDPIKENSFLPFOIDHROMAYOICQ 272
QY 237 SINFLEHNKLTHTDLKRENILFYKSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDE 296
D 273 SINFLEHNKLTHTDLKRENILFYKSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDE 332
QY 297 HNSTLVSTRHYRAPEVILALGWSOPCDVMSIGCILEYVGTVPFOHDSKELHAMMERI 356
D 333 HNSTLVSTRHYRAPEVILALGWSOPCDVMSIGCILEYVGTVPFOHDSKELHAMMERI 392
QY 357 LGPIPOHMIQKTRKRYFHHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416
D 393 LGPIPAHMIQKTRKRYFHHNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 452
QY 417 RRMLEVDPTORITLDEALQHPFDLLKKK 445
D 453 RRMLEVDPTORITLDEALQHPFDLLKKK 481

RESULT 2
CLKT_HUMAN STANDARD; PRT; 484 AA.
AC P49759;
AT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein kinase CLKT (EC 2.7.1.1-) (CLKT).
GN CLKT OR CLK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=91139618; PubMed=1704889;
RA Johnson K.W., Smith K.A.;
RT "Molecular cloning of a novel human cdc2/CD28-like protein kinase.";
RT J. Biol. Chem. 266:3402-3407(1991).
RN 12;
RP SEQUENCE FROM N.A.
RX MEDLINE=95082033; PubMed=7990150;
RA Hanes J.J., der Kameer H., Klaudiny J.J., Scheit K.H.;
RT "Characterization by cDNA cloning of two new human protein kinases.
RT Evidence by sequence comparison of a new family of mammalian protein
RT kinases.";
RT J. Mol. Biol. 244:665-672(1994).
CC -!- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;

```

```

CC CC IsoId=P49759-1; Sequence=Displayed;
CC CC Name=Short;
CC CC IsoId=P49759-2; Sequence=VSP_004852, VSP_004853;
CC CC Note=Lacks the kinase domain;
CC CC -!- PIM: Autophosphorylates on all three types of residue (By
CC CC similarity).
CC CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CC LAMMER SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
D 136 KRVCECIDHGMGMHVAVKIVKNGYRRAASEIOVLEHLNSTDPSNYSRCVOMLEWFDH 195
D 174 KRVCECIDHKGAGRHVAVKIVKNGVDCCEARSEIOVLEHLNSTDPSNYSRCVOMLEWFDH 233
QY 196 HSHVCIPELLGLSTYDPIKENSFLPFOIDHROMAYOICQSGINFLHNKLTHTDLKPEN 255
D 234 HSHVCIPELLGLSTYDPIKENSFLPFOIDHROMAYOICQSGINFLHNKLTHTDLKPEN 293
QY 256 ILFVSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDEHSTLVSTRHYRAPEVILA 315
D 294 ILFVSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDEHSTLVSTRHYRAPEVILA 353

```

QY	316	LGMSQPDWASICILILEYVLGFYQTHDSKEHLAMMERILGPIPOHMIQKTKRKYFH	373S
Db	354	LGMSQPDWASICILILEYVLGFYQTHDSKEHLAMMERILGPIPKMKIKTKRKYFH	4133
QY	376	HNOLDWDEHSSAGRYVRRRCCKPLKEFMLCHDEHEKLPFDLYARMLEYPDPTQITLDBALQ	435S
Db	414	HDRIDWDEHSSAGRYVRRRCCKPLKEFMLSQDVHERLFDLIQKMLEYPAKRITLREALK	4733
QY	436	HPEFDLTK	444
Db	474	HPEFDLTK	482
RESULT 3			
ID	CLK1_MOUSE	STANDARD;	PRT; 483 AA.
AC	P22518;		
DT	01-AUG-1991 (Rel. 19, Created)		
DT	01-AUG-1991 (Rel. 19, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Protein kinase CLK1 (EC 2.7.1.1-) (CLK) (Protein kinase SRY).		
GN	CLK1 OR CLK OR SRY.		
OS	Mus musculus (Mouse).		
OC	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91122038; PubMed=1825055;		
RA	Ben-David Y., Letwin K., Tannock L., Bernstein A., Pawson T.;		
RT	"A mammalian protein kinase with potential for serine/threonine and		
RT	tyrosine phosphorylation is related to cell cycle regulators.";		
RL	EMBO J. 10:317-325(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91094875; PubMed=1986248;		
RA	Howell B.W., Afar D.E., Lew J., Douville E.M., Icely P.L.,		
RA	Gray D.A., Bell J.C.;		
RT	"Sry, a tyrosine-phosphorylating enzyme with sequence homology to		
RT	serine/threonine kinases.";		
RL	Mol. Cell. Biol. 11:568-572(1991).		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).		
RX	MEDLINE=95394904; PubMed=7655564;		
RA	Duncan P.I., Howell B.W., Marais R.M., Drmanic S., Douville E.M.,		
RA	Bell J.C.;		
RT	"Alternative splicing of Sry, a nuclear dual specificity kinase.";		
RL	J. Biol. Chem. 270:21524-21531(1995).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).		
RX	MEDLINE=95082033; PubMed=7900150;		
RA	Hanes J.C., der Kammer H., Klaiding J.J., Scheit K.H.;		
RT	"Characterization by cDNA cloning of two new human protein kinases.		
RT	Evidence by sequence comparison of a new family of mammalian protein		
RT	kinases.";		
RL	J. Mol. Biol. 244:665-672(1994).		
RN	[5]		
RP	CHARACTERIZATION.		
RX	MEDLINE=96178222; PubMed=8617202;		
RA	Cowling K., Pawson T., Andrews B., Prasad J., Manley J.L., Bell J.C.,		
RA	Duncan P.I.;		
RT	"The Clk/Sty protein kinase phosphorylates SR splicing factors and		
RT	regulates their intranuclear distribution.";		
RL	EMBO J. 15:265-275(1996).		
CC	-1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS		
CC	OF THE SPLICESOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF		
CC	REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA		
CC	SPLICING. PHOSPHORYLATES SRINES, THREONINES AND TYROSINES.		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- ALTERNATIVE PRODUCTS:		
CC	Event=Alternative splicing; Named isoforms=2;		
CC	Name=Long;		
CC	Isoid=P22518-1; Sequence=Displayed;		

CC CC NameShort;
CC Iscd=P22518-2; Sequence=vsp_004654, vsp__004855;
CC Note=Lacks the Kinase domain;
CC -I PFM: Autophosphorylates on all three types of residue.
CC -II SIMILARITY: BELONGS TO THE SER/TRH FAMILY OF PROTEIN KINASES.
CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

CC EMBL, X57186, CAA40473.1, -;
DR DR EMBL, M38381, AAAA40151.1, -;
DR DR EMBL, U21209, AAC52257.1, -;
DR DR EMBL, I29221, AAG61485.1, -;
DR DR PIR, J43676, A33676,
DR PIR, I49275, I49275.
DR HSSP, Q00534, IBI8.
DR MGJ, MGI:107403, CLK.
DR InterPro: IPRO00719, Prot_Kinase.
DR InterPro: IPRO02290, Ser_thr_pkinase.
DR Pfam: PF00069, Pkinase, 1.
DR ProDom: PD000001, Prot_kinase, 1.
DR SMART: SMO0220, S_TKC_1.
DR PROSITE, PS00107, _PROTEIN_KINASE_ATP, 1.
DR PROSITE, PS00108, PROTEIN_KINASE_ST, 1.
DR PROSITE, PS50011, PROTEIN_KINASE_DOM, 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KV Tyrosine-protein kinase; Phosphorylation; Nuclear protein;
KW Alternative splicing.
KW KW
FT DOMAIN 29 33 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 160 476 PROTEIN KINASE.
FT NP_BIND 166 174 ATP (BY SIMILARITY).
FT BINDING 190 190 ATP (BY SIMILARITY).
FT ACT_SITE 287 287 BY SIMILARITY.
FT VASNPIC 130 135 KSHRK -> MKLLK (in isoform Short).
FT FT /FTid=vsp_004854.
FT VASNPIC 136 483 Missing (in isoform Short).
FT FT /FTid=vsp_004855.
FT MUTAGEN 190 190 K->R: LOSS OF ACTIVITY.
FT CONFLICT 379 379 P -> S (IN REF. 2).
FT CONFLICT 448 448 F -> L (IN REF. 2).
FT CONFLICT 453 456 VGKI -> IGKM (IN REF. 2).
SQ SEQUENCE 483 AA; 57095 MM; IDDEEC2AA2342619 CRC64;

Query Match 76.0%; Score 1832; DB 1; Length 483;
Best Local Similarity 79.7%; Pred.No.1.2e-109;
Matches 349; Conservative % 32; Mismatches 51; Indels 6; Gaps 4;

QY 307 YRAPEVILALGMSQPCDWSIGCIIIEYLGFTVQTHDSKEHLAMERIIIGPIPMHIQ 366
 DB 344 YRAPEVILALGMSQPCDWSIGCIIIEYLGFTVQTHDSKEHLAMERIIIGPIPMHIQ 403
 QY 367 KTRKREYHNNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPLVBRMLEYDPTQ 426
 DB 404 KTRKREYHNNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPLVBRMLEYDPTQ 463
 QY 427 RITLDEALQHPFDFLLKK 444
 DB 464 RITLKEALKHPFFYPLKK 481

RESULT 4

CLK2_MOUSE STANDARD; PRT; 499 AA.

ID CLK2_MOUSE STANDARD; PRT; 499 AA.
 AC 035491;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein kinase CLK2 (EC 2.7.1.-) (CDC-like kinase 2).
 GN CLK2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97439710; PubMed=9307018;
 RA Nayler O., Stamm S., Ullrich A.;
 RT "Characterization and comparison of four serine- and arginine-rich
 RT (SR) protein kinases.";
 RT Biochem. J. 326:693-700(1997).
 RL Biochem. J. 326:693-700(1997).
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- PTM: Autophosphorylates on all three types of residue.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: A003364; AAB87508.1; -.
 DR MGI: 1098669; CLK2.
 DR InterPro: IPR000719; Prot Kinase.
 DR InterPro: IPR002290; Ser Thr kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot kinase; 1.
 DR SMART: SM00220; S_TKc_1.
 DR SMART: PS00107; PROTEIN KINASE-ATP, 1.
 DR PROSITE: PS00108; PROTEIN KINASE-ST, 1.
 DR PROSITE: PS00111; PROTEIN KINASE-DM, 1.
 DR Transferrase: Serine/threonine-protein kinase; ATP-binding;
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein.
 FT DOMAIN 163 479
 FT NP_BIND 169 177 ATP (BY SIMILARITY).
 FT BINDING 192 192 ATP (BY SIMILARITY).
 FT ACT_SITE 289 289 BY SIMILARITY.
 SQ SEQUENCE 499 AA; 59955 MW; 092946BA171C9AMD CRC64;

Query Match 59.2%; Score 1427.5; DB 1; Length 499;
 Best Local Similarity 60.3%; Pred. No. 6.3e-84;
 Matches 261; Conservative 69; Mismatches 96; Indels 7; Gaps 3;

QY 18 SHYLEARS-----LNERDYRRRRYDEY-RNDYCEGYVRRYHRDIESGRRHCKSSVR 71

DB 50 SYHVSRSNDDSHSDRLRYRRRCGSRRRNDYSDRDEALAYDDDFRQSYEYHNSYSR 109
 QY 72 SRSS-PRKKNRHHSSHQSSKSHRRGRSIEDDEGHILICSGDYLRRARYEIVDTG 130
 DB 110 SQRSSRRRRRRRRRRSRTFSRSSSHSSRRKAXVEDDAAGHLIYHGDWLQRREYELVSLG 169
 QY 131 EGAFKAYVECTDHQMDGHNHVKIVKNVGRFRRARASTIQLYELHANSIDPNSVFCVOML 190
 DB 170 EGTSGRVQCVDRHGGRRVALKTIKVEEYKEARLEINLEKNEKPNKMLCVOMF 229
 QY 191 EWPFHGHVCIPELIGLSTYDFIKENSFLPEQIDIRQMAVOICOSINFLHNKLTPTD 250
 DB 230 DMFDYGHMCSFELLGISTFDPLKDNVYLPYPIHQVRHMAFOLCAVYKFLHDKLTHD 289
 QY 251 LKPNILFVYKSDYVYKNSKKKRDERTLKNIDIVVFGSATYDDEHHSITVSTRYAP 310
 DB 290 LKPNILFVYKSDYVYKNSKKKRDERTLKNIDIVVFGSATYDDEHHSITVSTRYAP 349
 QY 311 EVILALGMSQPCDWSIGCIIIEYLGFTVQTHDSKEHLAMERIIIGPIPMHIQ 370
 DB 350 EVILALGMSQPCDWSIGCIIIEYLGFTVQTHDSKEHLAMERIIIGPIPMHIQ 409
 QY 371 RKFYHNNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPLVBRMLEYDPTQ 430
 DB 410 RKFYHNNQLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPLVBRMLEYDPTQ 469
 QY 431 DEALQHPFDFLLKK 443
 DB 470 DEALQHPFDFLLKK 482

RESULT 5

CLK2_HUMAN STANDARD; PRT; 499 AA.

ID CLK2_HUMAN STANDARD; PRT; 499 AA.
 AC P49760; Q96C00;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein kinase CLK2 (EC 2.7.1.-) (CDC-like kinase 2).
 GN CLK2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95082033; PubMed=7990150;
 RA Hanes J.J., der Kammer H., Klaudiny J.J., Scheit K.H.;
 RT "Characterization by cDNA cloning of two new human protein kinases.
 RT Evidence by sequence comparison of a new family of mammalian protein
 RT kinases.";
 RT J. Mol. Biol. 244:665-672(1994).
 RL J. Mol. Biol. 244:665-672(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97474796; PubMed=9313372;
 RA Winfield S.L., Taveh N., Martin B.M., Gims E.I., Sidransky E.;
 RT "Identification of three additional genes contiguous to the
 RT glucocorticoidase locus on chromosome 1q21: implications for Gaucher
 RT disease.";
 RL Genome Res. 7:1020-1026(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong F.,
 RA Stajich M., Soares M.B., Bonaldi W.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein W.J., Ustin T.B., Toshitoki S., Carninci P., Prange C.,
 RA Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fandy U., Helton E., Keltman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley A.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Stein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=97066903; PubMed=8910305;
 RA Lee K., Du C., Horn M., Rabinow L.,
 RT "Activity and autophosphorylation of LAMER protein kinases.";
 RL J. Biol. Chem. 271:27299-27303(1996).
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P49760-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P49760-2; Sequence=VSP_004856, VSP_004857;
 CC -1- Note=lacks the kinase domain;
 CC -1- PTH: Autophosphorylates on all three types of residue.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L29218; AAA61482.1; -;
 DR EMBL: L29216; AAA61481.1; -;
 DR EMBL: AF023268; AAC51817.1; -;
 DR EMBL: BC014067; AAL14067.1; -;
 DR PIR: S53637; S53637.
 DR PIR: S53638; S53638.
 DR GeneW: HGNC:2069; CLK2.
 DR GK: P49760; -;
 DR MIM: 602989; -;
 DR GO: GO:0004674; P:protein serine/threonine kinase activity; TAS.
 DR InterPro: IPR000719; Prot. kinase.
 DR InterPro: IPR002290; Ser. Thr. kinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR ProDom: PD000001; Prot. kinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Tyrosinase; Serine/threonine-protein kinase; ATP-binding;
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein;
 KW Alternative splicing.
 FT DOMAIN 163 479 PROTEIN KINASE.
 FT NP_BIND 169 177 ATP (BY SIMILARITY).
 FT BINDING 193 193 ATP (BY SIMILARITY).
 FT ACT_SITE 290 290 BY SIMILARITY.
 FT VASPLIC 134 139 QHSRR->VSKSLAP (in isoform short).
 FT VASPLIC 140 499 /Ftd-VSP_004856.
 FT VASPLIC 140 499 Missing (in isoform short).
 FT VASPLIC 140 499 /Ftd-VSP_004857.
 FT CONFLICT 134 134 MISSING (IN REF. 3).

SEQUENCE 499 AA; 60090 MW; E43BFB3BAD6EE991 CRC64;
 Query Match 58.8%; Score 1417; DB 1; Length 499;
 Best Local Similarity 59.7%; Pred. No. 2.9e-83;
 Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;
 QY 18 SHYLEARS-----LNEPVDRRRYDVEY-RNDYCEGVPRHHRPIESGYRHCKSSVR 71
 DB 50 SYHVSRRSSYDSSDRRYDRYCGSYRRANDYSRDGDAYIDTDRHSYEVGRNSSYR 109
 QY 72 SRRSSPK-RKRNRRSSQHSRSKSHRRRSRIEDDEGHLICSGGVLRARVEIVDTL 129
 DB 110 SQRRSRKRRRRRRRRSRSPFRSSQHSRRRAKSVEDDAGHLIYHVGDMLOREYIVSTL 169
 QY 130 GEGARCKYVECTIDHGMDSNVAVKYKVVGRREARAEIQVLEHNTDPSVRCVM 189
 DB 170 GGTSTFRVQCVDRHRCGARVALKTIKVEKKEARAEIYNLEKINEKDPNNKVLCOVM 229
 QY 190 LEWFDHGHVCTVFELGLSTYDFTKENSFLPFOIDHROMAYQICQSGINPLHNKLTHT 249
 DB 230 FWFDFHGMCTSFELGLSTDFKDNNTLPIPIHQVRHMAFOLQAVKFLHDKLTHT 289
 QY 250 DKPENILFYKSDYVYKTKSKKRPDLKATDVDPFGSATYDDEHSTIVSTRHRA 309
 DB 290 DKPENILFYNSDIYELTYNLEKDRBSVSKTAIVAVVDFGSAFPHBHSITVSTRHRA 349
 QY 310 PEVILALGWSQPCDWSIGCILIEYLLGFTVFPQTHSKEHLAMERILGPIPOHMIOKTR 369
 DB 350 PEVILALGWSQPCDWSIGCILIEYLLGFTVFPQTHSKEHLAMERILGPIPSRIKTR 409
 QY 370 KRKYFHNDLWDHSSAGRYVRRCPLKEPMLCHDEHEKLPVLVRMLEYDPTQRT 429
 DB 410 KQYFPGRLDWDENTSAGRYVRNCKPLRRLYTSAEHHQLFDLISMLEYPAKRLT 469
 QY 430 LDEALQHPEDLLX 443
 DB 470 LGEALQHPFARLR 483
 RESULT 6
 CLK3 HUMAN STANDARD; PRT; 490 AA.
 AC P49761; O9BR53; O9BU77;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein kinase CLK3 (EC 2.7.1.1-) (CDC-like kinase 3).
 GN CLK3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=95082033; PubMed=7990150;
 RA Hanes J.J., der Kammer H., Klandi J.J., Scheit K.H.;
 RT "Characterization by cDNA cloning of two new human protein kinases.
 RT Evidence by sequence comparison of a new family of mammalian protein
 RT kinases.";
 RL J. Mol. Biol. 244:665-672(1994).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
 RC TISSUE=Cervix, and Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stajdolek M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Witting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield A.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marz M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1; Synonyms=long;
 CC Name=2; Synonyms=short;
 CC IsoId=P49761-1; Sequence=displayed;
 CC Note=Lacks the kinase domain;
 CC Name=3;
 CC IsoId=P49761-3; Sequence=VSP_004860;
 CC -1- PTM: Autophosphorylates on all three types of residue (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L29217; AAA61484.1; -;
 DR EMBL; L29220; AAA61483.1; -;
 DR EMBL; BC002555; AAH02555.1; -;
 DR EMBL; BC006103; AAH06103.1; -;
 DR EMBL; BC019881; AAH19881.1; -;
 DR PIR; S53639; S53639.
 DR PIR; S53640; S53640.
 DR HSSP; Q00534; 1B18.
 DR Genew; HGNC:2071; CLK3.
 DR GK; P49761; -;
 DR MTM; 602990; -;
 DR GO; GO:0004674; P;protein serine/threonine kinase activity; TAS.
 DR InterPro; IPR000719; Prot_Kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00693; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transfease; Serine/threonine-protein kinase; ATP-binding;
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein;
 KW alternative splicing.
 FT DOMAIN 156 472 PROTEIN KINASE.
 FT NP_BIND 162 170 ATP (BY SIMILARITY).
 FT BINDING 166 186 ATP (BY SIMILARITY).
 FT ACT_SITE 283 283 BY SIMILARITY.
 FT VARSPLIC 124 152 RSQSSKRSRSVEDDKGHVLCRIGDWI -> MELWGTWV
 FT VARSPLIC 153 490 /FTId=VSP_004858.
 FT VARSPLIC 158 180 /FTId=VSP_004859.
 FT VARSPLIC 158 180 /FTId=VSP_004860.
 FT VARSPLIC 158 180 /FTId=VSP_004860.

FT CONFLICT 132 133 SS -> TG (IN REF. 1).
 SQ SEQUENCE 490 AA; 58588 MW; 86A342AB8AB24CA CRC64;
 Query Match 54.8%; Score 1321; DB 1; Length 490;
 Best Local Similarity 55.3%; Pred. No. 3 6e-77;
 Matches 251; Conservative 60; Mismatches 101; Indels 42; Gaps 6;
 QY 4 PLEASHSVEEDTHPSHYLEARSINERDYDRDYDEYNDYCEGYVP-----R 51
 DB 43 PPRRSRSRSDRLP-----YORRYRERRRSDTYR---CERSPSPCEYDYGPSRSR 90
 QY 52 HYHDDISG-YRI-----HCKRSYSRRSRRSPKKNRHHSSQSSKSHRRRSRSD 105
 DB 91 HRRRSRERGYPYRKAHCHK-----RTRSCSSASSSQSSKSSRSVED 138
 QY 106 DEEGHLLCQSGDYLARAYEIVDTLGEAFGEKVECIDHGMDGMVAVKYKNGRYREA 165
 DB 139 DKEGHLCVCRIGDWLQERYEIVGNLGEQTFKVECDLHARGSQVALKIRANGKYREA 198
 QY 166 RSEIQLVHEHNSDPNSVFCVQMLFMPDHGHVCIYFELLGLSTYDFIKENSFLPQID 225
 DB 199 RLEINVAKKIKKDKENKFLCVMSDFNFGHMCIAPELLGKNTPEFLKNNPQYPLP 258
 QY 226 HIROMAYQICQSNFLHNHNLTHTDLPENILFVKSDYVVKYKSKKRDERTLKNTDIKY 285
 DB 259 HVRMAYQICHALRFLHENGULTHTDLKPEHILFVNSEPELTVNHHKSCSEKSVYNTSIRY 318
 QY 286 VDRGSATYDDEHNSLTVSTHYPAPVYLALGMSQCDVWSICGILIEYLGTVQTHD 345
 DB 319 ADFGSATFDEHHTTIVATRHYPPEVILLEGAPCDVWSIGCILEYRGGTTLFQTHE 378
 QY 346 SKEHLAMERILGPIDPMHIOKTRKRYFPHNOLDMDHSSARRCKPKPEMLCH 405
 DB 379 NRHLVMEKILGPIDPMHHRKRYFPGKGLVMDHSSDRYKYNENKPKSYTLQD 438
 QY 406 DEEHKLPDLYRMLEYDPTQRTILDALQHPFF 439
 DB 439 SLEHVQLDLMRRMLFEDPAORTTLBALHPFF 472
 RESULT 7
 CLK3 RAT STANDARD; PRT; 490 AA.
 AC Q63117;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein kinase CLK3 (EC 2.7.1.-) (CDC-like kinase 3).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-Sprague-Dawley, TISSUE=Brain.
 EX MEDLINE=96271481; PubMed=8679717;
 RA Becker W., Kentrup H., Henkelbach U., Joost H.G.;
 RT "cDNA cloning and characterization of rat CLK3, a LAMMER kinase
 RT predominantly expressed in testis.";
 RL Biochim. Biophys. Acta 1312:63-67(1996).
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- PTM: Autophosphorylates on all three types of residue.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X84351; CAA64076.1; --
CC PIR; S70352; S70352.
CC HSSP; Q00534; 1B18.

DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.

KW transferase; Serine/threonine-protein kinase; ATP-binding;
KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein.
FT DOMAIN 156 472 PROTEIN KINASE.
FT NP BIND 162 170 ATP (BY SIMILARITY).
FT BINDING 186 186 ATP (BY SIMILARITY).

FT ACT_SITE 283 283 BY SIMILARITY.
SQ SEQUENCE 490 AA; 58501 MW; 72F754F6464B19A CRC64;

Query Match 54.6%; Score 1317; DB 1; Length 490;
Best Local Similarity 55.1%; Pred. No. 6.6e-77;
Matches 250; Conservative 61; Mismatches 101; Indels 42; Gaps 6;

QY 4 PLEASHSVEDTHPHHYLEASLSNEDRYDEYNDCEGVV-----R 51
DB 43 PRRRSRSHDRIP-----YORRYREHDSOTYR--CEERSPSFGEDCYGSSRSR 90
QY 52 HYHRDIEEG-YRI-----HCSKSSVRSRRSSPKRNRHCSHQSRKSRKRSIED 105
DB 91 HRRRSRERGPYTRKHAHCHK-----RRTSSGSSASSRSGQSKRSRVED 138

QY 106 DEEGHLCQSGVLPARYEYDITLGEAGFGKVEYELIDGMGMVAIVKNGRYEAA 165
DB 139 DEEGHLCQSGVLPARYEYDITLGEAGFGKVEYELIDGMGMVAIVKNGRYEAA 198

QY 166 REEIOVLEHNTDENSVERCQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQID 225
DB 199 REEIOVLEHNTDENSVERCQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQID 258

QY 226 HIRQMAVQICQSLNHLNKLITDILKFNILFYKSDYVYKNSKMRDEKTLNDIKY 285
DB 259 HIRQMAVQICQSLNHLNKLITDILKFNILFYKSDYVYKNSKMRDEKTLNDIKY 318

QY 286 VDFGSATYDEHSHSLSTRYARAEVILALGMSQPCVWSIGCILLEYLYGFTVPTD 345
DB 319 VDFGSATYDEHSHSLSTRYARAEVILALGMSQPCVWSIGCILLEYLYGFTVPTD 378

QY 346 STEHLAMMERILGPIPOHYIOKTRKRYFHHNOLDWDEHSSAGYVRRCKPLKMLCH 405
DB 379 STEHLAMMERILGPIPOHYIOKTRKRYFHHNOLDWDEHSSAGYVRRCKPLKMLCH 438

QY 406 DEHEKFLPVLVRLMEYDPTORTLDDALQHPFF 439
DB 439 DEHEKFLPVLVRLMEYDPTORTLDDALQHPFF 472

RESULT 8
CLK3 MOUSE STANDARD; PRT; 490 AA.
AC 03432;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein kinase CLK3 (EC 2.7.1.-) (CDC-like kinase 3).
OS CLK3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97439710; PubMed=9307018;
RA Naylor O., Stamm S., Ullrich A.;
RT "Characterization and comparison of four serine- and arginine-rich
RT (SR) protein kinases.";
RL Biochem. J. 326:693-700(1997).

CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGinine-RICH (SR) PROTEINS
CC OF THE SPICEOSOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF
CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA
CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN TESTIS.
CC -1- PTM: Autophosphorylates on all three types of residue.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC LAMMER SUBFAMILY.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF033565; AAB87509.1; --
CC HSSP; Q00534; 1B18.
CC MGI; MGI:1098670; CLK3.

DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.

KW transferase; Serine/threonine-protein kinase; ATP-binding;
KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein.
FT DOMAIN 156 472 PROTEIN KINASE.
FT NP BIND 162 170 ATP (BY SIMILARITY).
FT BINDING 186 186 ATP (BY SIMILARITY).

FT ACT_SITE 283 283 BY SIMILARITY.
SQ SEQUENCE 490 AA; 58513 MW; A7C0311C530AC41 CRC64;

Query Match 54.4%; Score 1310; DB 1; Length 490;
Best Local Similarity 55.6%; Pred. No. 1.8e-76;
Matches 248; Conservative 62; Mismatches 110; Indels 26; Gaps 4;

QY 4 PLEASHSVEDTHPHHYLEASLSNEDRYDEYNDCEGVVPRHHRDIEGRI 63
DB 43 PRRRSRSHDRIP-----YORRYREHDSOTYR--CEERSPSFG--EDCYGS 86

QY 64 HCSKSSVRSRRSSPKRNRHCSHQSRKSRKRSIEDDEGHLIC 113
DB 87 HCSKSSVRSRRSSPKRNRHCSHQSRKSRKRSIEDDEGHLIC 146

QY 114 QSGDVLRAARYEYDITLGEAGFGKVEYELIDGMGMVAIVKNGRYEAAEIOYLE 173
DB 147 QSGDVLRAARYEYDITLGEAGFGKVEYELIDGMGMVAIVKNGRYEAAEIOYLE 206

QY 174 HLNSTDPSVFCVQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQIDIRQMAVQ 233
DB 207 HLNSTDPSVFCVQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQIDIRQMAVQ 266

QY 234 IQQSLNHLNKLITDILKFNILFYKSDYVYKNSKMRDEKTLNDIKYVDFGSATY 293
DB 267 IQQSLNHLNKLITDILKFNILFYKSDYVYKNSKMRDEKTLNDIKYVDFGSATY 326

QY 294 DDEHSTVSTRYARAEVILALGMSQPCVWSIGCILLEYLYGFTVPTDSEKHLAM 353
DB 327 DDEHSTVSTRYARAEVILALGMSQPCVWSIGCILLEYLYGFTVPTDSEKHLAM 386

QY 354 ERICGPIPOHNTQTRKRTFPHNOLDMDHSSAGRYVRRCRPLKPMCHDEHEKLP 413
 DB 387 EKILGPIPSHMTHTKQKTYFKGGLVMDNSSDGRVYKENCPLKSYMLQSLHEVOLF 446
 QY 414 DLVRMLEVDPRTITLDEALQHPFF 439
 DB 447 DLVRMLEVDPRTITLDEALQHPFF 472

RESULT 9
 DOA_DROME STANDARD; PRT; 517 AA.
 AC P49762: OGVAB8;
 DT 01-OCT-1986 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein Kinase Doa (Ec 2.7.1.-) (Protein darkener of apricot).
 GN DOA OR CG1658.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydriidea; Drosophilidae; Drosophila.
 CK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95011531; PubMed=7926721;
 RA Yun B., Farkas R., Lee K., Rabinow L.;
 RT "The Doa locus encodes a member of a new protein kinase family and is
 RT essential for eye and embryonic development in Drosophila
 RT melanogaster";
 RL Genes Dev. 8:1160-1173(1994).
 RN [2]
 RP REVISIONS.
 RA Rabinow L.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBD databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Berkeley;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champ M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaitai M., Kalish F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP FUNCTION, AND MUTAGENESIS OF LYS-199.
 RX MEDLINE=97066903; PubMed=8910305;
 RA Lee K., Du C., Horn M., Rabinow L.;
 RT "Activity and autophosphorylation of LAMMER protein kinases";
 RL J. Biol. Chem. 271:27299-27303(1996).
 CC -!- FUNCTION: Negative regulator of the copia retrotransposon element
 CC of the white (w) gene. In the eye, it is required for normal
 CC pigmentation, photoreceptor cell development and for organization
 CC of interommatidial reticles. Also essential for embryonic
 CC segmentation and differentiation of the nervous system.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT.
 CC -!- PFM: Autophosphorylates on all three types of residue.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X78715; CAA55367.1; -!
 DR EMBL: AE003767; AAF6832.2; ALT_INIT.
 DR PIR: A54099; A54099.
 DR FlyBase: FB0000480; Doa.
 DR GO: GO:0005737; C:cytoplasm; IDA.
 DR GO: GO:0005634; C:nucleus; IDA.
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IDA.
 DR GO: GO:0007350; P:blastoderm segmentation; IMP.
 DR GO: GO:0042051; P:eye photoreceptor development (sensu Drosop. . .); IEP.
 DR GO: GO:0007399; P:neurogenesis; IMP.
 DR GO: GO:0007548; P:sex differentiation; IGI.
 DR InterPro: IPR000719; prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; kinase_1.
 DR ProDom: PD000001; Prot_kinase_1.
 DR SMART: SM00220; S_TKc_1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP_1.
 DR PROSITE: PS00109; PROTEIN KINASE ST_1.
 DR PROSITE: PS00111; PROTEIN KINASE_DOM_1.
 KM Transferase; Serine/threonine-protein kinase; ATP-binding;
 KM Tyrosine-protein kinase; Phosphorylation.
 FT DOMAIN 170 484
 FT NP BIND 176 184 ATP (BY SIMILARITY).
 FT BINDING 199 199 ATP.
 FT ACT SITE 296 296 BY SIMILARITY.
 FT MUTAGEN 199 199 K->R: LOSS OF ACTIVITY.
 SQ SEQUENCE 517 AA; 60975 MW; 830EB8A50CEDA1 CRC64;

Query Match 50.7%; Score 1222; DB 1; Length 517;
 Best Local Similarity 57.8%; Pred. No. 7.6e-71;
 Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY 69 SVRRSSRRPKRRRHGSSHQ---SRKSHRRKRSR-----IE 104
 DB 92 SKRSRRSPASRRKHRYKRYDETSHSSRRRDRKDRDGRNRGQAKTAPVQ 151
 QY 105 DDERGHLICGSDVRLARYELVDTLGGAGCKVVECDHGDGMAVAKTVKNGRYREA 164
 DB 152 DDADGHLIYHGDTLHHRKYKMATLGEGTGRVVKVDMERD-YCALKILIIKVEKXREA 210
 QY 165 ARSELQVLEHNSDPSVRCQVLMFPHGHVCIIVFELLGLSTDFIKENSFLPQI 224
 DB 211 AKELINALEKIAQDPHCDHLCVMTIMWFYHGMCIVFEMLGLSVDFLRNNYEPYL 270

QY 225 DHIRQMAVQICQSIINFLHNKLTHTDLKPENLIFKYSYDVVKYNKMKRDELTAKTDTIK 284
 DB 271 DQVFNHAYQICYSKFLHNRRLHTDLKPENLIFVSDTSHNHNKINEVRYNTDVR 330
 QY 285 VVDGSAATYDEHNSHTVSTRHYRADEVILALGMSOPCVMSIGCTILIEYTGFTVFCOT 344
 DB 331 LIDGSAATPHEHNSHTVSTRHYRADEVILALGMSOPCVMSIGCTILIEYTGFTVFCOT 390
 QY 345 DSKENHLMERIIIGPDPQMTIQTRKRKIFHNOLDWDEHSSAGRYVRRCPLKEFMIC 404
 DB 391 DNEHLMERIIIGPDPQMTIQTRKRKIFHNOLDWDEHSSAGRYVRRCPLKEFMIC 449
 QY 405 HDEHEKLPDVRMLEYDPTORTITLDEALQHPEDLT 442
 DB 450 DSEHCHLPSLTKMELFEPSSRIITGEALHFFFDRL 487
 RESULT 10
 AF02 ARATH STANDARD; PRT; 427 AA.
 ID AF02 ARATH
 AC P51567;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein kinase AF02 (EC 2.7.1.-)
 GN AF02 OR AT4G24740 OR F22K18.60.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=95083650; PubMed=7991592;
 RA Bender J., Fink G.R.;
 RT "ARCT1, a LAMMER kinase from Arabidopsis thaliana, activates STE12-
 dependent processes in yeast.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Kuromori T., Yamamoto M.;
 RT "A thaliana gene encoding protein kinases of a new family.";
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schaeffler C., Mamput R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Ertan K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Grievell L.A., Rieger M., Mueller M.,
 RA Weichselgartner W., de Simone V., Obermaier B., Macho R., Mueller M.,
 RA Kreis W., Delseny M., Puigdomenech P., Watson M., Schmidtkehl T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hobeisel U., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandebussche F.,
 RA Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzsaecker T., Bothe G., Rampsberger U., Hilbert R., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Woolfjan P., Klein Lanckhorst R., Rose M., Hauf J., Koester P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Giesen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,
 RA Petter A., Rajandream M.A., Lyne M., Benes V., Reimann S.,
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan W., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Aiguiton A., Vitale D., Lignocri R., Piravandi E.,
 RA Messner O., Ungley F., Clapaud G., Muendlein A., Feilner R.,
 RA Schnabl S., Hiltner R., Schmidt W., Lecharny A., Aubourg S.,

RA Chedder F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandembol M., Bagnus M., Terol J., Torres A.,
 RA Perez-Perez A., Punelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Hejlskov L., Schwarz S., Scholler P., Heber S., Francs P., Bielle C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Beyan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Pannell L., Dedbia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sehon M., Murray J., Sheet P., Cordes M., Abu-Threideh U.,
 RA Stoneking T., Kallick J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Mux P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kremer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spiech J., Ryan E., Andrews S., Giesel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drome K., Cotton M., Joshi C.,
 RA Antonou B., Zidan M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Wa P., Zhong J., Preston R., Vill D., Shekhar M., Matro A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodi M., Johnson A.,
 RA Chen B., Marra M., Martienssen R., McCombie W.R.;
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RT Nature 402:769-777(1999).
 RL Nature 402:769-777(1999).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U16177; AAA57118.1; -;
 DR EMBL; D45353; BAA08214.1; -;
 DR EMBL; AL035356; CA822989.1; -;
 DR EMBL; AL161562; CA879384.1; -;
 DR PIR; T05560; T05560.
 DR HSSP; P24941; HCL.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR002290; Ser Thr kinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR ProDom; PD000001; Prot kinase; 1.
 DR SMART; SM00220; S_TKC_1; PROTEIN KINASE ATP, FALSE NEG.
 DR PROSITE; PS00107; PROTEIN KINASE ATP, FALSE NEG.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 1.
 DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 98 423
 FT NP BIND 104 112 ATP (BY SIMILARITY).
 FT BINDING 127 127 ATP (BY SIMILARITY).
 FT ACT_SITE 223 223 BY SIMILARITY.
 SQ SEQUENCE 427 AA; 49566 MW; 33A1948B8372B374 CRC64;
 Query Match 35.2%; Score 848.5; DB 1; Length 427;
 Best Local Similarity 48.1%; Pred. No. 3.5e-47;
 Matches 167; Conservative 70; Mismatches 97; Indels 13; Gaps 8;
 QY 104 EDEEGLILQSGDIVLFAREIVDTIGEGAFGRVCEIDHMDGHAAYAKYKVRKYE 163
 DB 79 EDDXGHIYELGDDTLPRYKIKSKGEGFGVLCWBRERKM-VAVKIVGVKRYE 137
 QY 164 AARSEIOLVHLNSTDPNSVFCVOMLEMDHGHGIVVELGLSTYFIKENSFLPFO 223
 DB 138 AAMEIEMLQGLGHDGCG-NRCVQIRNMVDYNNHICIFEKKGSSLYFLRKNVRSFP 196
 QY 224 IDHRCQMAVQICQSIINFLHNKLTHTDLKPENLIFKYSYDV-VK-NSMKRD---ERT 277
 DB 197 IDVREIGWGLTCVAFMDLRIHIDLKPENLIVSSDYVKIPEYKSGRLQRDVCYKEV 256
 QY 278 LKNDIVVFGSATVDDEHNSHTVSTRHYRADEVILALGMSOPCVMSIGCTILIEYTG 337
 DB 257 PKSSAIVDFGSGTTRRQDQTYIVSTRHYRADEVILALGMSOPCVMSIGCTILIEYTG 316

QY 338 FVTFQTHDSKELAMMERLGPIDPMIOTK--RKKYFHHNOLDWDEHSSAGRYRRRC 395
 Db 317 EALFQTHNLEHLMAMERVLGPIDPMIOTK--RKKYFHHNOLDWDEHSSAGRYRRRC 376
 QY 396 K--PLKEFPMICH-DEEHEKLPDVRMLLEYDPTORTLDEALQHPFF 439
 Db 377 KLPRLONLIMOHVDSAGELIMVQGLRFPDSERITAREALRHPPF 423

RESULT 11
 AFCL ARATH STANDARD; PRT; 467 AA.
 AC P51566: Q39184; (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Protein kinase AFCL (EC 2.7.1.-).
 GN AFCL OR AME2 OR ATG53570 OR PAPI2.270.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucotyledon; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN 1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=95083650; PubMed=7991592;
 RA Bender J., Fink G.R.;
 RT "AFCL, a LAMMER kinase from Arabidopsis thaliana, activates STE12-
 RT dependent processes in yeast."
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).
 RN 2)
 RP SEQUENCE FROM N.A.
 RA Kurumori T., Yamamoto M.;
 RT "A thaliana genes encoding protein kinases of a new family."
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RN 3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=1130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansgore W., Unseld M.,
 RA Farmann M., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Deleney M., Boutry M., Grievell L.A., Maché R., Nigamanch P.,
 RA De Simone V., Choise N., Artiguenave F., Robert C., Broclet P.,
 RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quelier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmach E., Drzonek H., Erle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Bangert P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppe S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Joehner T.-H., Nordiek G.,
 RA Reichelt U., Scharte M., Schen O., Barques M., Terol J., Clément J.,
 RA Navarro P., Collado C., Perez-Perez A., Oltgenwelder B., Duchemin D.,
 RA Coore R., Lande M., Berger-Liauro C., Purnelle B., Masny D.,
 RA De Haan M., Maere A.C., Alcaraz J.-P., Coutet A., Casacuberta E.,
 RA Montfort A., Argitrou A., Flores J., Liguori R., Vitale D.,
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mees H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Uteirack T., Fujii C.Y., Shea T.P.,
 RA Cressy T., Miltscher J., Sellers P., Gill J.E., Feldblum T.V.,
 RA Pail G., Miltscher J., Sellers P., Gill J.E., Feldblum T.V.,
 RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kohara M., Matsumoto M., Matsushima K., Kishida Y.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Matsumae A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:820-822(2000).
 CC -1- FUNCTION: ACTIVATOR OF YEAST TRANSCRIPTION FACTOR, STE12.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U16176; AAA57117.1; -
 DR EMBL: D45354; BAA08215.1; -
 DR EMBL: AL132966; CAB67664.1; -
 DR PIR: S71169; S71169.
 DR HSBP: P24941; 1A01.
 DR InterPro: IPR000719; Prot kinase.
 DR InterPro: IPR002290; Ser thr kinase.
 DR InterPro: IPR001245; Tyr kinase.
 DR Pfam: PF00069; kinase; 1.
 DR ProDom: PD000001; Prot kinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR SMART: SM00219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP, FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 115 443 PROTEIN_KINASE.
 FT NP BIND 121 129 ATP (BY SIMILARITY).
 FT BINDING 144 144 ATP (BY SIMILARITY).
 FT ACT_SITE 240 240 BY SIMILARITY.
 FT CONFLICT 117 117 I -> M (IN REF. 1).
 SQ SEQUENCE 467 AA; 54198 MW; A885FD32CE11B181 CRC64;

Query Match 34.4%; Score 830; DB 1; Length 467;
 Best Local Similarity 45.8%; Pred. No. 5.7e-46;
 Matches 160; Conservative 71; Mismatches 102; Indels 16; Gaps 8;

QY 105 DDEEGHLCQSGGVLPARYEIVDTLGEAGFQVCEICDHGMQMVAVKIVKNGRYREA 164
 Db 97 DDKDGHVFEVGGTLLPRYQILSKMGEGTFCGLCEFDNKNREV-VAIKVIRINKYRBA 155
 QY 165 ARSEIQLVLEHNTSDNSVFRQVOMLEFDDHGHVIVELGLSTYDICKENSLFPOI 224
 Db 156 AMEIDVLQRLTRHDVG-SRCVQIRNMFYRNHITVEKLGPSLYDLRKNYSFPI 214
 QY 225 DTRQMAVOICQISINFLHNKLTHTDLKPELIFVKSQV---VKYNSKMRDEKTLN 280
 Db 215 DLRKLGRLQLESVAVMHDLRIHDLKPELILVSSSEYIKIPDYKFLSRPTKDSYFNX 274
 QY 281 ----TQIKVDFRSATYDDHSHSTVSTHYRPAEVLALGMSQPCDWSIGCLILEYL 336
 Db 275 LPRSSAIKILDFSTTFEHDHNYVSTHYRPAEVLILGVGMVPCDLSIGCLILELCS 334
 QY 337 GFVTFQTHDSKELAMMERLGPIDPMIOTK--RKKYFHHN-OLDWDEHSSAGRYRR 393
 Db 335 GEAFLPQTHNLEHLMAMERVLGPIDPMIOTK--RKKYFHHN-OLDWDEHSSAGRYRR 394
 QY 394 --RCKPLKEFPMICH-DEEHEKLPDVRMLLEYDPTORTLDEALQHPFF 439
 Db 395 VWKLPLRLONLIMOHVDSAGELIMVQGLRFPDSERITAREALRHPPF 443

RESULT 12
 AFCL ARATH STANDARD; PRT; 400 AA.
 AC P51568: Q39185; (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein kinase AFCL (EC 2.7.1.-).
 GN AFCL OR AME2 OR ATG53266 OR PAPI1.140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

CC eucoside II; Brassicales; Brassicaceae; Arabidopsis.
 CC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kurumori T., Yamamoto M.;
 RT "A thaliana gene encoding protein kinases of a new family.";
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Meyer K.F.X., Schueler C., Wandt R., Murphy G., Volkhaert G.,
 RA Pohl T., Duesterhoef A., Strikema W., Entian K.-D., Terryn N.,
 RA Harris B., Anstoege W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Portelle D., Perez-Alonso M., Boutry M., Schmitthelm T.,
 RA Vos P., Hobeisel U., Zimmermann W., Medler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandebussche F.,
 RA Breken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weltzenegger T., Bothe G., Rampsberger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirke W.,
 RA Wooljan P., Klein Lanhorst R., Rose M., Haut J., Koetter P.,
 RA Berneser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gieles J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
 RA Petter A., Rajendram M.A., Lyne M., Benes V., Reckman T.-H.,
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehner S.,
 RA Dose S., de Haan M., Maarse A.C., Schafer M., Mueller-Auer S.,
 RA Gabel C., Fuchs W., Fartmann B., Grandtath K., Dauner D., Hezel A.,
 RA Neumann S., Argitlon A., Vitale D., Liguori R., Plavanti E.,
 RA Massenot O., Quigley F., Cladaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheffor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Hejman L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stoeck S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Pannell L., Dehila N., Gnoj L., Schutz K., Huang E., Siegel L.,
 RA Sehon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Steneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Lattelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Mixx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Matis E., Dante W., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Sheker M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohay N., Hasegawa A., Hamed A., Lohi M., Johnson A.,
 RA Chen E., Marra M., Matlensen R., McCormie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 RN [3]
 RP SEQUENCE OF 6-400 FROM N.A.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=95083650; PubMed=7991592;
 RA Bender J., Fink G.R.;
 RT "AFL1, a LAMMER kinase from Arabidopsis thaliana, activates STE12-
 dependent processes in yeast.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D45355; BAA08216.1; -
 DR EMBL; AL022537; CAI18595.1; -
 DR EMBL; AL161581; CAB79983.1; -
 DR EMBL; U16178; AAS57119.1; -
 DR PIR; T04460; T04460.
 DR HSSB; P24941; IAO1.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKC_1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 71 396
 FT NP BIND 77 85
 FT BINDING 100 100
 FT ACT SITE 196 196
 FT COMPACT 8 8
 FT SEQUENCE 400 AA; 46224 MW; 8F707ABDCFB1352 CRC64;
 Query Match 32.8%; Score 790; DB 1; Length 400;
 Best Local Similarity 41.7%; Pred. No. 1,7e-43;
 Matches 169; Conservative 70; Mismatches 134; Indels 32; Gaps 11;
 QY 57 IESGRIHCKSSVRS-----RSPPKR--NRHSGHQSRSKRRRSSTED 105
 DB 2 IANGFE-SMDKRRKRPMTWDEAPAPPAKRAVKGHSDRIISPUR-----DD 53
 QY 106 DEEGHLCQSGDVARAYEIVDILGEGAFKVEGICDHGMDGHWVAKIYKNGRYREA 165
 DB 54 DRDHVYFSURKNDLTPRYKILSKMGEGTFRVLECMDBDK- EYVAIKIRSKKRDAA 112
 QY 166 RSEIVLEHLNSTDPNSVFCVOMLEWPDHGHVCIVFELLSTGYDFIKNSFLPPQID 225
 DB 113 MIEDIVLOKLVKSKDKRT-RCVOMKWFEDYRNHICIVFEKLGSLDFELKKNKSAFPLA 171
 QY 226 HIRGMAVOIGSINFLHNHNLTPDILKPEVILFVKSQDVYVYKSKKMBDRT-----LKN 280
 DB 172 LVRFPGGLLESAVYHMLDVLVTDUKPEVILVSEENVKLPNKKSANETFRCLPKS 231
 QY 281 TDIVVDFGSATYDDEHNSITLVSTRYARAEVITLALGWSQPCDVWSIGCILIEYVIGFTV 340
 DB 232 SAKILDFGSITVCNRIHHSIVOTRYHRSPEVILIGWSYQCDLWSIGCILFELCTGEAL 291
 QY 341 FQTHDSKEHLAMERILGPIPOHMICKTR--RKVFHN--QLMDHSSAGRYVR--RRC 395
 DB 292 FQTHDNEHLAMERILGPIDEHTRKASFGAKRYRRCGRCLMPPGANSRESIRAVKRL 351
 QY 396 KPLKFMILCH-DEHEKFLPLVRMLLEYDFTORITLDEALQHPFF 439
 DB 352 DRLKDWVSKHVDNTRSRFADLLVGLLAYDPSERLTANEDLDRHF 396
 RESULT 13
 LKHL SCHPO STANDARD; PRT; 575 AA.
 ID LKHL SCHPO
 AC 010156; 09C319;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Protein kinase lkh1 (EC 2.7.1.37) (EC 2.7.1.112).
 OS LKHL OR SPAC104.11C.
 GN Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 RN [1]

RP SEQUENCE FROM N.A., AND FUNCTION.
 RX MEDLINE=2161699; PubMed=11741326;
 RA Kim K.-H., Cho Y.-W., Kang W.-H., Kim J.-H., Byun K.-H., Park Y.-D.,
 RA Bae K.-S., Park H.-M.;
 RT "Negative regulation of filamentous growth and flocculation by Lknl,
 RT a fission yeast LAMMER kinase homolog.";
 RL Biochem. Biophys. Res. Commun. 289:1237-1242(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Best N., Hayes J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., Melean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
 RA Weltyens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.M.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler M., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleaux V., Motier S.,
 RA Galibert F., Ayres S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: Protein kinase that may act as a negative regulator of
 CC filamentous growth and flocculation.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- PTM: Autophosphorylates on all three types of residue (by
 CC similarity). BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC LAMMER SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 CC EMBL: AF34941; AAK12335.1;
 DR EMBL: 269239; CAD29835.1;
 DR PIR: JC7794; JC7794.
 DR HSSP: Q00534; IB18.
 DR GeneDB: SPACID4.11c;
 DR InterPro: IPR000719; Prot_Kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_Kinase; 1.
 DR SMART: SM00220; S_TKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1;
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding;
 KW Tyrosine-protein kinase; Phosphorylation.
 FT DOMAIN 17 118 PRO-RICH.
 FT DOMAIN 247 567 PROTEIN_KINASE.
 FT NP_BIND 253 261 ATP (BY SIMILARITY).

FT BINDING 276 276 ATP (BY SIMILARITY).
 FT ACT_SITE 373 373 BY SIMILARITY.
 SQ SEQUENCE 575 AA; 65022 MW; 62E0C600EC35839A CRC64;
 Query Match 31.8%; Score 767; DB 1; Length 575;
 Best local Similarity 41.5%; Pred. No. 7.3e-42;
 Matches 141; Conservative 78; Mismatches 115; Indels 6; Gaps 3;
 QY 105 DDEBGLICQSGDVLBARIEYDVLCEGAFKQVVEICDGMGMVAVKIVNNGYREA 164
 Db DDDGHYKVVNSKFNRYTVVRLGHGFGKVIQCCYDGS-TGRHCATVTRAIPIYREA 287
 QY 165 ARSEIOVLEHNTDSTNSVRCQMLEPDHGHVCEIVELGLSYDPIKENSFLPFOI 224
 Db SLEIRLVQTIASDPTNEKCIQLMDYRKHICITVDLFGMSVFDLKNNNYIPFL 347
 QY 225 DTRQNAVQICQSIENFLHNKLTHTDLPKENILFVKSVDYVVKNSPKKDETLKNTDK 284
 Db KHIQMSQQLFKSVAFHSLGLVHTDLPKENVILVNASRTIRLPYRNSQKVIANSCEIR 407
 QY 285 VVPGSGATYDDEHSTLVSTRHARAEVILAGMSQPCDWSIGCILVEYLGFYQTH 344
 Db LIDFGSATFEDVHSSVSTRHARAEVILGLGMSYPCDWSIGCILVEYLGFYQTH 467
 QY 345 DSKELAMMERLIGPIPOHMI---QTRKRYFHNQDLPDEHSSAGRYVR--RCKPK 399
 Db EDEHLCMEKILGPDNRNIMSRSSRTSQRFFKSDKVRYPISNTPKKSINYLQSLTLE 527
 QY 400 EPMLCDEHEKFLPDLVRKLEVDLPQRTITDGLQHPFF 439
 Db QIRAVSSPEVALLDLILDKRVFVDPKRRITAKALMHPFF 567
 RESULT 14
 ID KNS1_YEAST STANDARD; PRT; 737 AA.
 AC P23350; Q12399;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Protein kinase KNS1 (EC 2.7.1.37) (EC 2.7.1.112).
 GN KNS1 OR YIL019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=91375437; PubMed=1910150;
 RA Padmanabha R., Gehring S., Snyder M.;
 RT "The KNS1 gene of Saccharomyces cerevisiae encodes a nonessential
 RT protein kinase homologous that is distantly related to members of the
 RT CDC28/cdc2 gene family.";
 RL Mol. Gen. Genet. 229:1-9(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=S288C; PubMed=9046100;
 RA Purnelle B., Goffeau A.;
 RT "The sequence of 32kb on the left arm of yeast chromosome XII reveals
 RT six known genes, a new member of the ser/pauperins family and a new
 RT AAS transporter homologous to the human multidrug resistance
 RT protein.";
 RL Yeast 13:183-188(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313267; PubMed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansgorge W.,
 RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
 RA Ertian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,

RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Porteleile D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Uristarazu L.A., Vandenbol M., Verhaaselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RA Nature 387:87-90(1997).
RN [4]
RP FUNCTION.
RX MEDLINE=97066903; PubMed=8910305;
RA Lee K., Du C., Horn M., Rabinow L.;
RT "Activity and autophosphorylation of LAMMER protein kinases.";
RU J. Biol. Chem. 271:27299-27303(1996).
CC [5]
CC AUTOPHOSPHORYLATION.
RX MEDLINE=21616999; PubMed=11741326;
RA Kim K.-H., Cho Y.-M., Kang W.-H., Kim J.-H., Byun K.-H., Park Y.-D.,
RA Bae K.-S., Park H.-M.;
RT "Negative regulation of filamentous growth and flocculation by Lhl1,
RT a fission yeast LAMMER kinase homolog";
RU Biochem. Biophys. Res. Commun. 289:1237-1242(2001).
CC [1]
CC BIOCHEM. NONESSENTIAL PROTEIN KINASE.
CC [1]
CC CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC [1]
CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC [1]
CC PTM: Autophosphorylates on all three types of residue.
CC [1]
CC SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC LAMMER SUBFAMILY.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>,
CC or send an email to license@isb-sib.ch).
CC [1]
CC EMBL: M85200; AAA34724.1; -
DR EMBL: 273125; CA97468.1; -
DR EMBL: X97560; CA96171.1; -
DR EMBL: 273123; CA97465.1; -
DR PIR: S64767; S64767.
DR SGD: S0003942; KMS1.
DR InterPro: IPR000719; Prot. Kinase.
DR InterPro: IPR002290; Ser_Thr_PKinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Prot_Kinase; 2.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00108; PROTEIN KINASE ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Transferrase; Serine/threonine-protein kinase; ATP-binding;
KM Tyrosine-protein kinase; Phosphorylation.
FT DOMAIN 143 146 POLY-SER.
FT DOMAIN 148 152 POLY-ASN.
FT DOMAIN 313 720 PROTEIN KINASE.
FT NP BIND 319 327 ATP (BY SIMILARITY).
FT BINDING 343 343 ATP (BY SIMILARITY).
FT ACT SITE 440 440 BY SIMILARITY.
FT DOMAIN 673 679 SER/THR-RICH.
FT CONFLICT 411 421 RFPSSHQAIA -> GSPALISGHC (IN REF. 1).
SQ SEQUENCE 737 AA; 83842 MW; F0224445BDEB8431 CRC64;

Query Match 29.8%; Score 718.5; DB 1; Length 737;
Best Local Similarity 33.3%; Pred. No. 1.2e-38;
Matches 156; Conservative 90; Mismatches 123; Indels 99; Gaps 11;

QY 67 KSVYRSRRSPKRRNRNHCSSHQSRKSRKRSRTIEDDEGHLLCQSGDVLR--RYE 124
DB 261 QSNLKQIGSLRFTSGSSESAS-----NKSNEKTKDKGHVYQENDIDGSGGRV 314

QY 125 IVDTLGGAGREKVECTIDHGMDCMNVKTVNNGVRRAARSEIOVLEHNTDPSNVF 184
DB 315 VKDLLGQCFKFKVLCIDNKYEPYVAVKIRADRYREAKTELRLTLLNNDPQGF 374
QY 185 RCVOALEWFDHGHVACIVFELLGLSTVDFPKENSFLPFQIDHROMAYQICQSFILHN 244
DB 375 QCLLRBEFDKKNHICLVTDLYGSRISYDFPMCSNGIANFPSSHQAIALRQIRVCFHDL 434
QY 245 KLHTLDKPEHILVKSQDV-----VKNSMKRD-----ERTKATDIKVPDGA 291
DB 435 GIHTDLKPEHILLCDETHIAQKPLKTVOSLSKRREARSGKRIKPEIKIIFGSA 494
QY 292 TYDEHSHSTVSTHYRAPEVIALGMSOPDWSICILIEYVLGFTVQDHSKHLA 351
DB 495 IFHYEHPVSTHYRAPEVIALGMSOPDWSICILIEYVLGFTVQDHSKHLA 554
QY 352 MWERILG-PIPOHNIQKTRK-----KYPHNN-OLDWDEHSSAG-- 388
DB 555 MMQRINGTPPTDILDKMFYKSKHLGNSPSDLNSTVIAKHFDKRTLSLOWPEKVRGDTI 614
QY 389 -----RYVRRRC-----KPLKEPMLCH----- 405
DB 615 TTESMKRVLOSCLRDIYISKVLDQYGSLSINMLPPEPKMSLSKSLAMKRGTHSS 674
QY 406 -----DEBEKLF-----DLVRMLVPTQRTITLDEALQHPFDL 441
DB 675 SSSTDELKKEPFLFWYFIDLLEKMFEPPTKRITAKDLDEHMFVL 722

RESULT 15
DYR2_HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.1).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eimbler K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
RT specificity of DYRK-related kinases, a novel family of dual
RT specificity protein kinases.";
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
RA Diatchenko L., Marinina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stadleren M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosk S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalón D.K., Wuzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzyzanski M.I., Skaleka U., Smalhus D.E.,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:15:04 ; Search time 40 Seconds

(without alignments)
2870.835 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410
Sequence: 1 MCIPLASHSVEDTHPSHY.....QRITLDALQHPFDLKKK 445

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2315.5	96.1	481	4	Q9HAZ1
2	1893	78.5	484	4	Q8NSV8
3	1879	78.0	484	6	Q8BGC2
4	1803	74.8	453	4	Q2NRU6
5	1607	66.7	301	11	Q99JL6
6	1604	66.6	301	11	Q8CEU9
7	1437	59.6	498	11	Q91YR2
8	1305	54.1	490	11	Q8CIV1
9	1252	52.0	341	4	Q8BRG8
10	1222	50.7	511	5	Q8IMM0
11	1222	50.7	580	5	Q85RC9
12	1180	49.0	832	5	Q8T041
13	965.5	40.1	903	5	Q17917
14	963	40.0	409	5	Q8MY22
15	839.5	34.8	219	11	Q35721
16	839	34.8	431	10	Q49967

17	802	33.3	434	10	Q943M7	Q943M7 oryza sativ
18	789	32.7	437	10	Q9M598	Q9M598 mesembryant
19	778	32.3	430	10	Q94UC8	Q94UC8 oryza sativ
20	648.5	26.9	881	5	Q9BKX8	Q9BKX8 plasmodium
21	648.5	26.9	881	5	Q8IL19	Q8IL19 plasmodium
22	598	24.8	1300	3	Q8X0V5	Q8X0V5 neurospora
23	550.5	22.8	538	11	Q8BM34	Q8BM34 mus musculu
24	550.5	22.8	586	11	Q922X0	Q922X0 mus musculu
25	549.5	22.8	601	4	Q9BRB5	Q9BRB5 homo sapien
26	546.5	22.7	474	4	Q8WS54	Q8WS54 oikopleura
27	544.5	22.6	568	4	Q9HB76	Q9HB76 homo sapien
28	544.5	22.6	588	4	Q9HB77	Q9HB77 homo sapien
29	532.5	22.1	722	5	Q9V3D5	Q9V3D5 drosophila
30	524	21.7	632	11	Q8B155	Q8B155 mus musculu
31	523	21.7	508	5	Q20604	Q20604 caenorhabdi
32	523	21.7	796	5	Q9TVF4	Q9TVF4 caenorhabdi
33	523	21.7	817	5	Q9XTF3	Q9XTF3 caenorhabdi
34	522	21.7	461	5	Q8T2V5	Q8T2V5 trypanosoma
35	517	21.5	520	4	Q8NEF2	Q8NEF2 homo sapien
36	509	21.1	607	5	Q917R8	Q917R8 drosophila
37	509	21.1	908	5	Q9VX07	Q9VX07 drosophila
38	495.5	20.6	828	5	Q8MT39	Q8MT39 drosophila
39	483.5	20.1	629	11	Q9EPW2	Q9EPW2 mus musculu
40	480	19.9	608	5	Q8WQ88	Q8WQ88 entamoeba h
41	477	19.8	1157	10	Q9FX39	Q9FX39 arabidopsis
42	474	19.7	882	5	Q8WOL7	Q8WOL7 caenorhabdi
43	474	19.7	948	5	Q22155	Q22155 caenorhabdi
44	472	19.6	1138	10	Q8RX85	Q8RX85 arabidopsis
45	472	19.6	1266	10	Q9LSH3	Q9LSH3 arabidopsis

ALIGNMENTS

RESULT 1

Q9HAZ1 PRELIMINARY; PRT; 481 AA.

AC Q9HAZ1, 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Protein serine/threonine kinase CLK4.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21100912; PubMed=11170754;
RA Schultz J., Jones T., Bork P., Sheer D., Blencke S., Steyrer S.,
RT Wellbrock U., Beyec D., Ullrich A., Wallasch C.,
RT "molecular characterization of a cDNA encoding Functional Human CLK4
kinase and localization to Chromosome 4q35.";
RL Genomics 71:368-370(2001).
RN (2)
RP SEQUENCE FROM N.A.
RA Schultz J., Jones T., Bork P., Sheer D., Blencke S., Steyrer S.,
RT Wellbrock U., Beyec D., Ullrich A., Wallasch C.,
CC Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF294429; AAG10074.1; --
DR HSSP; Q00534; 1B18.
DR GeneW; HGNC:13659; CLK4.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_Kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 481 AA; 57492 MW; F402C3685CDA306 CRC64;

Query Match 96.1%; Score 2315.5; DB 4; Length 481;
 Best Local Similarity 96.2%; Pred. No. 1,9e-186;
 Matches 432; Conservative 2; Mismatches 4; Indels 11; Gaps 2;

QY 8 SH-SVEEDTH-----PSHYLEARSINERDRIYDEYNDYCEGYPRRYHND 56
 Db 33 SHSTOENRICKPHHQKESDCHYLEARSLNERDYDRYVDEYNDYCEGYPRRYHND 92
 QY 57 IESEYRIHCKSSVRSRSPKRRKRNHCSHOSRKSRRKRSRSIEDDEGHILCOQG 116
 Db 93 IESEYRIHCKSSVRSRSPKRRKRNHCSHOSRKSRRKRSRSIEDDEGHILCOQG 152
 QY 117 DVLRAREIVDTLGEAGFGVVEICDHGMGMVAKIVNVRGEARESEIQTLEHIN 176
 Db 153 DVLRAREIVDTLGEAGFGVVEICDHGMGMVAKIVNVRGEARESEIQTLEHIN 212
 QY 177 STDPNSVRCVOMLEWFDHGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQ 236
 Db 213 STDPNSVRCVOMLEWFDHGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQ 272
 QY 237 SINPLHNKLTHTDLKPEMLFVKSDYVYKSKMRDERLTKNIDIKVDFGSATYDDE 296
 Db 273 SINPLHNKLTHTDLKPEMLFVKSDYVYKSKMRDERLTKNIDIKVDFGSATYDDE 332
 QY 297 HNSTLVSTRHYRAPEVILALGWSQPCDWSIGCILLEYLGFVPTQTHSKELAMMERI 356
 Db 333 HNSTLVSTRHYRAPEVILALGWSQPCDWSIGCILLEYLGFVPTQTHSKELAMMERI 392
 QY 357 LGPILPQHMIOKTRKRYFHNOLDMDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLV 416
 Db 393 LGPILPQHMIOKTRKRYFHNOLDMDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLV 452
 QY 417 RMLEVDPTQRTITLDEALQHPFDLLKK 445
 Db 453 RMLEVDPTQRTITLDEALQHPFDLLKK 481

RESULT 2

Q8NSV8 PRELIMINARY; PRT; 484 AA.
 ID Q8NSV8;
 AC Q8NSV8;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE CXC-like kinase 1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=bone;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC031549; AA31549.1;
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Transferase.
 SQ SEQUENCE 484 AA; 57290 MW; F34B5B44988BD118 CRC64;

Query Match 78.5%; Score 1893; DB 4; Length 484;
 Best Local Similarity 82.3%; Pred. No. 6.8e-151;
 Matches 353; Conservative 30; Mismatches 44; Indels 2; Gaps 2;

QY 18 SHYLEARSINERDRIYDEYNDYCEGYPRRYHNDIESGRTHCSKSSVRSRSP 77
 Db 54 SHYLEARSINERDRIYDEYNDYCEGYPRRYHNDIESGRTHCSKSSVRSRSP 113
 QY 78 KRR-RNRHCSH--OSRKSRRKRSRSIEDDEGHILCOGSDVLRAREIVDTLGEAGFG 135
 Db 114 KSKRIHSHSTRSHRSRSHKSRKRRKTRSEVDEBHLCOGSDVLSAREIVDTLGEAGFG 173
 QY 136 KVEICIDHGMGMVAKIVNVRGEARESEIQTLEHINSTDPNSVRCVOMLEWFDH 195
 Db 174 KVEICIDHGMGMVAKIVNVRGEARESEIQTLEHINSTDPNSVRCVOMLEWFDH 233
 QY 196 HGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQINFLHNKLTHTDLKPE 255
 Db 234 HGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQINFLHNKLTHTDLKPE 293
 QY 256 ILFVKSDYVYKSKMRDERLTKNIDIKVDFGSATYDDEHNSTLVSTRHYRAPEVILA 315
 Db 294 ILFVKSDYVYKSKMRDERLTKNIDIKVDFGSATYDDEHNSTLVSTRHYRAPEVILA 353
 QY 316 LGWSQPCDWSIGCILLEYLGFVPTQTHSKELAMMERILGPQHMIOKTRKRYFH 375
 Db 354 LGWSQPCDWSIGCILLEYLGFVPTQTHSKELAMMERILGPQHMIOKTRKRYFH 413
 QY 376 HNOLDMDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLEVDPTQRTITLDEALQ 435
 Db 414 HDRLDMDDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLEVDPTQRTITLDEALQ 473
 QY 436 HPPFDLLKK 444
 Db 474 HPPFDLLKK 482

RESULT 3

Q9BGQ1 PRELIMINARY; PRT; 484 AA.
 ID Q9BGQ1;
 AC Q9BGQ1;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical 57.3 kDa protein.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 OC Cercopitheidae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Frontal cortex;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB056421; BAB33079.1;
 DR HSSP; O00534; 1818
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Hypothetical protein; ATP-binding; Kinase;
 KW Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 484 AA; 57318 MW; F5324D4007119878 CRC64;

Query Match 78.0%; Score 1879; DB 6; Length 484;
 Best Local Similarity 79.3%; Pred. No. 1e-149;
 Matches 356; Conservative 30; Mismatches 51; Indels 12; Gaps 4;

QY 8 SHSVEED-----TH-----PSHYLARSINERDYPDRRYVDEYRNDYCEGVYPRYHRDI 57
 DB 34 SHHSARBNKRCYTHYTHKNCDSHTLSEKINENEDYHSRRYIDYRNDYNQCEPGRHRDH 93
 QY 58 ESGYRIHCSKSVSRSSPKRK-RNRCSH-QSRKSHRKRKRSIEDDEGHLCQS 115
 DB 94 EERYQHSKSGSGRSGRSYKSKRH IHTSHRSRSHGSHRKRKTRVEDDEBHHLCQS 153
 QY 116 GYVLLARYEIVDTLDEGAFGRVYECIDHGMGMVAAYIVKNGRYRMAASEIQTUEHL 175
 DB 154 GYVLLARYEIVDALBGAFAFKVYECIDHKAQRHAAVIVKNGRYCEAARSEIQTUEHL 213
 QY 176 NSTDPNSVRCVQMLEWEDHGHVCIIVELGLSTYDPIKENSFLPFOIDIRPMAVOIC 235
 DB 214 NTDNNSYTRCYQMLEWEDHGHVCIIVELGLSTYDPIKENGFLPRLDHRMAVOIC 273
 QY 236 QSINFLHNKLTHTDLKPENILFKVSDYVVKYNSKMKDEKTLKNTDIKVDFGSATYDD 295
 DB 274 KSVNPLHNSKLTHTDLKPENILFKVSDYVVKYNSKMKDEKTLKNTDIKVDFGSATYDD 333
 QY 296 EHSHTLVSTRHYRADEVILALGMSOPCDVMSIGCLIEYVGLFTVFTQTHSKHLLAMMER 355
 DB 334 EHSHTLVSTRHYRADEVILALGMSOPCDVMSIGCLIEYVGLFTVFTQTHSKHLLAMMER 393
 QY 356 ILGPIPOHMIQKTRRKYFHHNOLDDEHSSAGRYVRRCPLKEFMLCHDEHEKLPDL 415
 DB 394 ILGPIPKHMIQKTRRKYFHHNOLDDEHSSAGRYVRRCPLKEFMLCHDEHEKLPDL 453
 QY 416 VRRMLEYDPTORTLDEALQHPFDLLKK 444
 DB 454 IQKMLEYDPAKITLKEALKHPFDLLKK 482

RESULT 4
 ID Q9NRL6 PRELIMINARY; PRT; 453 AA.
 AC Q9NRL6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CLK4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=bone marrow;
 RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;
 RT "A novel gene expressed in human bone marrow";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AF212224; AAF87326.1; --
 DR HSSP: Q00534; 1B18.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00207; S_TKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 453 AA; 53806 MW; 3A2AC6A3FFAE9461 CRC64;

Query Match 74.8%; Score 1803; DB 4; Length 453;
 Best Local Similarity 80.4%; Pred. No. 2,4e-143;
 Matches 344; Conservative 16; Mismatches 40; Indels 28; Gaps 3;

QY 18 SHYLLARSINERDYPDRRYVDEYRNDYCEGVYPRYHRDIHSGRIHCSKSVSRSSP 77
 DB 54 SHYLLARSINERDYPDRRYVDEYRNDYCEGVYPRYHRDIHSGRIHCSKSVSRSSP 113

QY 78 KKRNRHCSHOSRSKSRKRKRSIENDEBHHLCQSGDYLRARYEIVDTLDEGAFGRV 137
 DB 114 KSKRHIF-----HSTSHRSH-----GD-----EIVDTLDEGAFGRV 145
 QY 138 VECIDHGMGMVAAYIVKNGRYRMAASEIQTUEHLNSTDPNSVRCVQMLEWEDHGH 197
 DB 146 VECIDHKAQRHAAVIVKNGRYCEAARSEIQTUEHLNTDIPNSTPFCVQMLEWEDHGH 205
 QY 198 HVCYIFELLGLSTYPIKENSFLPFOIDIRPMAVOICQSNFLHNKLTHTDLKPENIL 257
 DB 206 HVCYIFELLGLSTYPIKENGFLPRLDHRMAVOICQSNFLHNKLTHTDLKPENIL 265
 QY 258 FKSDPVVYVNSKMKDEKTLKNTDIKVDFGSATYDDEHSTLVSTRHYRADEVILALG 317
 DB 266 FKSDPVVYVNSKMKDEKTLKNTDIKVDFGSATYDDEHSTLVSTRHYRADEVILALG 325
 QY 318 WQOPCDVMSIGCLIEYVGLFTVFTQTHSKHLLAMMERILGPIPOHMIQKTRRKYFHHN 377
 DB 326 WQOPCDVMSIGCLIEYVGLFTVFTQTHSKHLLAMMERILGPIPOHMIQKTRRKYFHHN 385
 QY 378 QLDMDHSSAGRYVRRCPLKEFMLCHDEHEKLPDLVRRMLEYDPTORTLDEALQHP 437
 DB 386 QLDMDHSSAGRYVRRCPLKEFMLCHDEHEKLPDLVRRMLEYDPTORTLDEALQHP 445
 QY 438 FFDLLKKK 445
 DB 446 FFDLLKKK 453

RESULT 5
 ID Q99JL6 PRELIMINARY; PRT; 301 AA.
 AC Q99JL6;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to CDC like kinase 4.
 GN CLK4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: BC002220; AAH02220.1; --
 DR MGI: MGI:1098551; CLK4.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 301 AA; 35768 MW; 7F157E2E28C6C1 CRC64;

Query Match 66.7%; Score 1607; DB 11; Length 301;
 Best Local Similarity 98.3%; Pred. No. 4,4e-127;
 Matches 296; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 145 MDMGMVAAYIVKNGRYRMAASEIQTUEHLNSTDPNSVRCVQMLEWEDHGHVCIYFE 204
 DB 1 MDMGMVAAYIVKNGRYRMAASEIQTUEHLNSTDPNSVRCVQMLEWEDHGHVCIYFE 60
 QY 205 ILGLSTYDPIKENSFLPFOIDIRPMAVOICQSNFLHNKLTHTDLKPENILFKVSDYV 264
 DB 61 ILGLSTYDPIKENSFLPFOIDIRPMAVOICQSNFLHNKLTHTDLKPENILFKVSDYV 120
 QY 265 VYVNSKMKDEKTLKNTDIKVDFGSATYDDEHSTLVSTRHYRADEVILALGMSOPCDV 324
 DB 121 VYVNSKMKDEKTLKNTDIKVDFGSATYDDEHSTLVSTRHYRADEVILALGMSOPCDV 180

QY 325 WSICILIEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRKRYPHNQLDMDHE 384
 DB 181 WSICILIEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRKRYPHNQLDMDHE 240
 QY 385 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLK 444
 DB 241 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLK 300
 QY 445 K 445
 DB 301 K 301

RESULT 6

Q8CEU9 PRELIMINARY; PRT; 301 AA.
 AC Q8CEU9; 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CDC like kinase 4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RA MEDLINE=22354683; PubMed=12466851;
 RA THE FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL: AK013974; BAC25420.1;
 SQ SEQUENCE 301 AA; 35740 MW; 58EADD0DD228C6DE CRC64;

Query Match 66.6%; Score 1604; DB 11; Length 301;
 Best Local Similarity 98.0%; Pred. No. 7.8e-127;
 Matches 295; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 145 MDGHWAVKIVKNGRYREARASEIOYLEHNLSTDPNSVRCVOMLEWEPDHGVCIVE 204
 DB 1 MDGHWAVKIVKNGRYREARASEIOYLEHNLSTDPNSVRCVOMLEWEPDHGVCIVE 60
 QY 205 LLSSTYDFIKENSFLPFOIDHROMAYQICOSINFHNKLTHTDLKPENILFVKSDDY 264
 DB 61 LLSSTYDFIKENSFLPFOIDHROMAYQICOSINFHNKLTHTDLKPENILFVKSDDY 120
 QY 265 VKVSKMKRDERTLKNTDIKVDGSAFYDDEHSTLSTRHYRAPVILALGMSQPCV 324
 DB 121 VKVSKMKRDERTLKNTDIKVDGSAFYDDEHSTLSTRHYRAPVILALGMSQPCV 180
 QY 325 WSICILIEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRKRYPHNQLDMDHE 384
 DB 181 WSICILIEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRKRYPHNQLDMDHE 240
 QY 385 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLK 444
 DB 241 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLK 300
 QY 445 K 445
 DB 301 K 301

RESULT 7

Q91YR2 PRELIMINARY; PRT; 498 AA.
 AC Q91YR2; 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CDC-like kinase 2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strubeberg R.,
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC015080; AA015080.1;
 DR InterPro: IPR001064; Crystalin.
 DR InterPro: IPR000719; Prot kinase.
 DR InterPro: IPR02290; Ser thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR SMART: SM00220; S_trc; 1.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; kinase; Transferase.
 SQ SEQUENCE 498 AA; 59956 MW; 0A87B74D78889B19 CRC64;

Query Match 59.6%; Score 1437; DB 11; Length 498;
 Best Local Similarity 60.6%; Pred. No. 1.7e-112;
 Matches 262; Conservative 69; Mismatches 95; Indels 6; Gaps 3;

QY 18 SHYLEARL---NRDVRDRRYDEY-RNICYEGYVRNHRHRLDSGRYHCKSSVRS 72
 DB 50 SHYRSRSDYDHSRRYDRYCGSYRNYSRDRGEAYDTDFRQSYEYHRENSYSRS 109
 QY 73 RRS-PRKRNHCSHOSRSHRRKRSIEDDEGHLFCOSGDVLRAREYIVDTLGE 131
 DB 110 QRSSRRKRRRRRRRTSRSSSHSRRAKVEDDAEHLHYVQDMQERYEIVSTIGE 169
 QY 132 GAFGVVECIDHGMDCMAVAVKIVKNGRYREARASEIOYLEHNLSTDPNSVRCVOMLE 191
 DB 170 GTFGRVVOCVDRRSGTVALKIIKYVEKYEAARLEINVEKINEKDPENKNCVQWFD 229
 QY 192 WFDHGHVCIYELIGSTYDFIKENSFLPFOIDHROMAYQICOSINFHNKLTHTDL 251
 DB 230 WFDYHGHVCIYELIGSTYDFIKENSFLPFOIDHROMAYQICOSINFHNKLTHTDL 289
 QY 232 KEENILFYKSDYVVKYKMKRDERTLKNTDIKVDGSAFYDDEHSTLSTRHYRAPE 311
 DB 290 KEENILFYKSDYVVKYKMKRDERTLKNTDIKVDGSAFYDDEHSTLSTRHYRAPE 349
 QY 312 VILALGMSQPCVWMSIGCILEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRK 371
 DB 350 VILALGMSQPCVWMSIGCILEYLGFTVQTHDSKEHLAMERILGPIQHMIOKTRK 409
 QY 372 KYFHNQLDMDHESSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLD 431
 DB 410 KYFYGRLDMDHESSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLD 469
 QY 432 EALQHPFDLLK 443
 DB 470 EALQHPFDLLK 461

RESULT 8

Q8CIV1 PRELIMINARY; PRT; 490 AA.
 AC Q8CIV1; 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CDC-like kinase 3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK090215; BAC41138.1;
 SQ SEQUENCE 490 AA; 58469 MW; 2B2B71B9A3DA1D2 CRC64;

 Query Match 54.1%; Score 1305; DB 11; Length 490;
 Best Local Similarity 55.4%; Pred. No. 2.1e-101;
 Matches 247; Conservative 62; Mismatches 111; Indels 26; Gaps 4;

 QY 4 PLEASHVEDTHPSHYLLEKSLNERDYDRRYVDYKNDYCEGYPRHYHDISSGRI 63
 DB 43 PPRSRSRSHDRIP-----YQRRYRHRSDTYR---CEERSPSFG---EDCYGS 86
 QY 64 HCSKSVSRSSRSPKR-----KRNHCSHQSRSKSHRRKRSIEDDEGHILIC 113
 DB 87 SRSHRRSRREPAPRYRRKKAHHCHKRRTRSCSSASRSQSSKRSRVEDDKGHILVC 146
 QY 114 QSGDVLARVEIVDTLGEAFGKYVVCIDHGMGMHVAKYKNVGRYPARSETIQLVE 173
 DB 147 RIGDMLOERYEIVGNIGEGTFGKVCCLDARGKSGVALKIRNVGKYREARLEINVLK 206
 QY 174 HLNSTDNSVFCOMLEMFHDHGHVCIYFELLGLSTYDFIKENSFLPQIDHIRMAYQ 233
 DB 207 KIKKDKENKFLCVLMSDMWTFHGMCIAPFELLGKNTPEFLKNNQPLPLHYHMAVQ 266
 QY 234 IQCSINFLHNKLTHTDLKPNILFYKSDYVVKYNSKMRDELTNDIKYVDFGSATY 293
 DB 267 LCHALRFLHNNOLAHDTLKPENILFYVSEFETLYNHNKCEKSVNNTSIRVADGSAIF 326
 QY 294 DDEHSTVLTSTRYARAEVILALGWSQPCDWSIGCILEYVLGFVPGTDSKHEHLM 353
 DB 327 DHEHTITVATRYRPRPEVILBLGMAQPCDWSIGCILEYVIRGFTLPHTHNRHLVYM 386
 QY 354 ERILGPIPMIIOKTRRKRYFHHNQLDWDHSSAGRYVRRCKPLKEMLCHDEHEKLF 413
 DB 387 EKILGPIPSMIRHTRKQKFKYKGLVWDENSDGYVKNCKPLKSYMLQSLHVOVF 446
 QY 414 DLVARMLEYPDRTITLDEALQHPFF 439
 DB 447 DLVARMLEYPDRTITLDEALQHPFF 472

 RESULT 9
 Q9BRG8 PRELIMINARY; PRT; 341 AA.
 ID Q9BRG8;
 AC Q9BRG8;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DB Similar to CDC-like kinase 2 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA Strauberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBS databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; BC006274; AA06274.1;
 DR InterPro: IPR001064; Crystalin.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Prodom: PD000001; Prot_kinase; 1.

DR SMART: SM00220; 5_TKC; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMA; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 FT NON TER 1
 SQ SEQUENCE 341 AA; 40236 MW; AF43DA22455A3D4 CRC64;

 Query Match 52.0%; Score 1252; DB 4; Length 341;
 Best Local Similarity 67.7%; Pred. No. 3.9e-97;
 Matches 218; Conservative 53; Mismatches 51; Indels 0; Gaps 0;

 QY 122 RYELVDLGGAGKGVCEICDHGMGMHVAKYKNVGRYPARSETIQLVEHNSITDP 181
 DB 4 RYELVDLGGAGKGVCEICDHGMGMHVAKYKNVGRYPARSETIQLVEHNSITDP 63
 QY 182 SVPRCVQMLWPFHGHVCIYFELLGLSTYDFIKENSFLPQIDHIRMAYQIQSINFL 241
 DB 64 NKNLCVQMPWDFHGHVCIYFELLGLSTYDFIKENNYLPYPIHVRHMAFQCAVFL 123
 QY 242 HNNKLTHTDLKPNILFYKSDYVVKYNSKMRDELTNDIKYVDFGSATYDEHNSIT 301
 DB 124 HNNKLTHTDLKPNILFYKSDYVVKYNSKMRDELTNDIKYVDFGSATYDEHNSIT 183
 QY 302 VSTRHYARAEVILALGWSQPCDWSIGCILEYVLGFVPGTDSKHEHLMERILGPI 361
 DB 184 VSTRHYARAEVILALGWSQPCDWSIGCILEYVLGFVPGTDSKHEHLMERILGPI 243
 QY 362 QHMIQKTRRKRYFHHNQLDWDHSSAGRYVRRCKPLKEMLCHDEHEKLFVLVRRM 421
 DB 244 SRMIRKTRKQKYPFRGLDWDHSSAGRYVRRCKPLRLYITSEAEHHQLFOLISM 303
 QY 422 YDPTQRTITLDEALQHPFFDLK 443
 DB 304 YEPARKLTIGDALQHPFFARLR 325

 RESULT 10
 Q8IMMO PRELIMINARY; PRT; 511 AA.
 ID Q8IMMO;
 AC Q8IMMO;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE CG1658-PB.
 GN DOA.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballweir L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.R., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Doolan K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ibegam C.,
 RA Talati M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,
 RA Kimmel B.E., Kodira C.D., Kratz C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meriklov G., Malshe N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paele J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spieding A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkars R., Tecor C., Turner C., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett M., Wolley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195 (2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Goecky J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnak D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegam C., Jalali M., Kruse D., Li P., Matei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
 RA Pacieb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svrtkars R., Tecor C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome,"
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seattle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.,
 RT "Annotation of *Drosophila melanogaster* genome,"
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AB003767; AAF56833.2;
 SQ SEQUENCE 511 AA; 60210 MW; 806366BFBDB8A948 CRC64;

Query Match 50.7%; Score 1222; DB 5; Length 511;
 Best Local Similarity 57.8%; Pred. No. 2.2e-94;
 Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY SVSRSSSPKRRKRRKSSHQ---SRSKRRKRRS-----IE 104
 DB 86 SRSHRKSPPASRRQHKRYRDETHSSSRHRDRADDERDSSGRNNRQSAKTAAPVIG 145
 QY DDEBHLTGOSGDVLRARVEYDITGEGAFGVVECDIHGMGMHVAIVKNGRYREA 164
 DB 146 DDADGHLVHTGDIHHRKYMATIGEGTFGVVVKVMDERD-YCMALKIINVKRYREA 204
 QY ARSEIQLVLEHNSLTPNSVFRVQVLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224
 DB 165 ARSEIQLVLEHNSLTPNSVFRVQVLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224

DB 205 AKLEINALEKIAQKQDCHQKLCVKMIDWPDYHGMCIIVEMGLSVDFLRNNYEPYL 264
 QY 225 DHIRQNAVYOICOSINFLHNKLTHTDLPKENTLPFKSDYVVKYNSKMRDETLKNTDVK 284
 DB 265 DQVRHNAAYOCISVSKFLHNRLTHETDLKPEKNTLPVDSYTSYHNKINREVARVKTQVDR 324
 QY 285 VVPGSATVDEHHSHTVSTRHRAPEVITLACMSQPCVMSIGCTLFEYLGFWFOCH 344
 DB 325 LIDPGSATVDEHHSHTVSTRHRAPEVITLACMSQPCVMSIGCTLFEYLGFWFOCH 384
 QY 345 DSKHEHMMERLIGPIPOHMIQTRKRKYHHNOLDMDHESAGRYVRRCRPLKEFMVC 404
 DB 385 DNEHLMAMERLIGQIPYMARKT-KTKFYHGKLDMDKSSAGRYVRRCRPLKEFMVC 443
 QY 405 HDEHEKLPDLVRKMLEYPTORTIDEALQHPFDLL 442
 DB 444 DSEDHCELFSLIKKMLEYPPSSRTITGELHHPFDRL 481

RESULT 11
 ID 095RC9 PRELIMINARY; PRT; 580 AA.
 AC 095RC9;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE LD44053p.
 GN DOA OR CG1658.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neuroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 NC NCBI TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacieb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY061474; AAL29022.1; -
 DR FlyBase; FBgn0000480; Doa.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc_1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 580 AA; 67894 MW; F73AC5B47733AF3A CRC64;

Query Match 50.7%; Score 1222; DB 5; Length 580;
 Best Local Similarity 57.8%; Pred. No. 2.5e-94;
 Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY SVSRSSSPKRRKRRKSSHQ---SRSKRRKRRS-----IE 104
 DB 155 SRSHRKSPPASRRQHKRYRDETHSSSRHRDRADDERDSSGRNNRQSAKTAAPVIG 214
 QY 105 DDEBHLTGOSGDVLRARVEYDITGEGAFGVVECDIHGMGMHVAIVKNGRYREA 164
 DB 215 DDADGHLVHTGDIHHRKYMATIGEGTFGVVVKVMDERD-YCMALKIINVKRYREA 273
 QY 165 ARSEIQLVLEHNSLTPNSVFRVQVLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224
 DB 274 AKLEINALEKIAQKQDCHQKLCVKMIDWPDYHGMCIIVEMGLSVDFLRNNYEPYL 333
 QY 225 DHIRQNAVYOICOSINFLHNKLTHTDLPKENTLPFKSDYVVKYNSKMRDETLKNTDVK 284

Db 334 DQVHMAVQCYSVKELHNDRLHTLDPKPNILFVSDYTSYHNKINEVRKNTDVR 393
Qy 285 VVPGSGATYDEHSTSTVSTRHYRAPEVIALGMSQPCVWSIGCLILEYIGFTYFOTR 344
Db 394 LIDPGATFHEHSTSTVSTRHYRAPEVIALGMSQPCVWSIGCLILEYIGFTYFOTR 453
Qy 345 DSKEHLAMERILGIPPOHNIQKTRKKYFHHNOLDMEHSSAGRYVRRCRPLKEFMLC 404
Db 454 DNEBHLAMERILGIPPOHNIQKTRKKYFHHNOLDMEHSSAGRYVRRCRPLKEFMLC 512
Qy 405 HDEBHEKLPDVRKMLFYDPTORTLDEALQHPFOLL 442
Db 513 DSEBHEKLPDVRKMLFYDPTORTLDEALQHPFOLL 550

RESULT 12
Q8T041 PRELIMINARY; PRT; 832 AA.
AC Q8T041;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE L03161P (CG1658-PC).
GN DOA OR CG1658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclet J., Paragas V., Park S., Phoumenavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandel M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Boyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Bailly J.F., Agbayani A., An H.J., Andrews-Piankocch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Bouchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hestlin D., Houston K.A., Howland T.J., Wei M.H., Iobagwan C.,
RA Urali M., Kalush F., Karpen G.H., Ke Z., Kennison J.G., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Styriak R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RA Science 287:2185-2195 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC Celinker S.E., Adams M.D., Krommler B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hestlin D., Howland T.J.,
RA Iobagwan C., Jallai M., Kruse D., Li P., Matvei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Styriak R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome";
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Mitra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bengman C., Berman B., Carlson J.W., Celinker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommler B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seattle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome";
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069573; AAL39718.1; -
DR EMBL; AF003767; AAN14305.1; -
DR FlyBase; FBgn000480; Doa.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR Prodom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 832 AA; 95723 MW; 6EC90CB1A031D68 CRC64;
Query Match 49.0%; Score 1180; DB 5; Length 832;
Best Local Similarity 53.8%; Pred. No. 1,4e-99;
Matches 222; Conservative 65; Mismatches 94; Indels 32; Gaps 4;
Qy 61 YRIHCKSVSRSSPRKRNKRCSSHQSS-----KSHRRKRS----- 102
Db 391 YASPASSSSNNKQPQPPQQQQQQQSQSQGSNSVFNHSGQHPHQQQQNMMSYALGLH 450
Qy 103 -----IENDEGHILICQSGDVLRAKRYELVDTLGGAGAGKVVCEICDHGDSGNHVAVKI 154
Db 451 FETAKRPVTDADAGHLIYHGDILHHRKYIMATLGGTGRVVKVMDERD-YOMALKI 509
Qy 155 VKNGRYREARSEIQVLEHLNSTDPNSVRCVOMLEWPFHGHGVCIVFELIGSTYDFI 214
Db 510 IKVKEKREARKETLNLEKIAQNDPCHDLGVGMIDPFGHGMICVPEMLIGISVDFL 569
Qy 215 KENSFLPQIDHITQMAVQCYQSGINFLHANKLHTLDPKPNILFVSDYTSYHNKINEVRKNTDVR 274

```

Db      570 REINNYFPYLDQVRHNAVQLCYSAKFLHNNRLTHTDLPENILFVDSQYTHNKKIRE 629
Qy      275 ERLTKRTDKVDFGSAITDDHNSITVSTRHYRAPVILALGNSQPCDWSIGCILLEY 334
Db      630 VRRVKNQTVRLIDFSGATFDHSHSTIVSTRHYRAPEVILELGMWQPCDWSIGCILFEL 689
Qy      335 YLGFVTFQTHDSKEHLAMMERILGPIPI-----QHMIQKTRKKYHHNQLDMDDESSAGR 389
Db      690 YIGITLTFQTHDRHFLAMMERILGQIPYMAANNHILSKTKTKYTHGKLDMDDESSAGR 749
Qy      390 YVRRRCKPLKEFMLCHDEBEHKLFDVVRMLEYDPTQRTITLDEALQHPFDLL 442
Db      750 YVRDHCKPLFLCQLSDPSDHCFLSLIKKMLEYEPSSRITLGEALHHPFDRL 802

```

RESULT 13

```

Q17917 PRELIMINARY; PRT; 903 AA.
ID Q17917; Q19042;
AC Q17917; Q19042;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E02H4.3 protein.
GN E02H4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Aliscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favejello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kistler N., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray J., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Souhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spoat J., Womlam P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Barlow K.;

```

```

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; Z70205; CAA94122.1; -.
DR EMBL; Z68003; CAA94122.1; JOINED.
DR EMBL; Z68003; CAA91979.1; -.
DR EMBL; Z70205; CAA91979.1; JOINED.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002965; P_tich_extensn.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; PKINASE; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 903 AA; 98863 MW; 3EB78CE79C9DA187 CRC64;

```

Query Match 40.1%; Score 965.5; DB 5; Length 903;
Best Local Similarity 46.6%; Pred. No. 1.7e-72;

```

Matches 185; Conservative 72; Mismatches 99; Indels 41; Gaps 6;
Qy      66 SKSSVSRSSRSPK-----RKRNRHSSQSSSKSRKRKRSTEDDEGHILCOGSDYL 119
Db      507 SRSGLOASQARPPVPEIVSSQRTQ-----QDDKQGHLYSKGDFI 547
Qy      120 RAREIVDTLGEAGAFGVCEIDHGMGNVAVIKVNGRYEARSEIQVLEHNSD 179
Db      548 IARFTYDILQSGTGFQGVNR-VNDSLSDTFPMALKIKVNSKYEAALKEKVLQKLAED 606
Qy      180 PNVSVRCQMLENPFDDHGHVCIFFELIGLSTYDFIKENSFLPQDHIROMAYQCSIN 239
Db      607 PEKKNWVHNGSYFDVNGHICLFLDMSGSIPEFLKANKYPYMEQTLHITQLCNAV 666
Qy      240 FLHNNKLTHTDLPENILFVDSQYTHNKKIR-----MKRDE-----RTLKNTDIXV 285
Db      667 FLHNNKLTHTDLPENILFVDSQYTHNKKIRVLRFRKIDDCQPLRVHSHVRL 726
Qy      286 VDFGSAITDDHNSITVSTRHYRAPVILALGNSQPCDWSIGCILLEYLGFVQTHD 345
Db      727 IDFGSATFDHSHSTIVSTRHYRAPEVILELGMWQPCDWSIGCILLEYLTVGTLFQTHE 786
Qy      346 SKHEHAMMERILGPIQHMIOKTRKKYHHNQLDMDDESSAGRYYRRCKPLKEFYLCH 405
Db      787 NREHLAMMERVLDIDIPRVAKRT-KTFPIINGRLDWVNTSADAAVVRDCKPLRISVCT 845
Qy      406 DEBEHKLFDVVRMLEYDPTQRTITLDEALQHPFDLL 442
Db      846 DPEHVELFELIEMLMMEPLARKKLPALQHPHFRNL 882

```

RESULT 14

```

Q08MLY2 PRELIMINARY; PRT; 409 AA.
ID Q08MLY2;
AC Q08MLY2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E02H4.3b protein.
GN E02H4.3 OR E02H4.3B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sims M.A.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=99069613; PubMed=9851916;
RX none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Barlow K.;

```

```

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z70205; CAD44096.1; -.
DR EMBL; Z68003; CAD44096.1; JOINED.
DR EMBL; Z68003; CAD44105.1; -.
DR EMBL; Z70205; CAD44105.1; JOINED.
DR WormPep; E02H4.3b; CB31463.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; PKINASE; 1.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyTc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.

```

DR PROSITE; PS00108; PROTEIN_KINASE_ST, 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 409 AA; 47383 MW; F13B36A9CBA97AE CRC64;

Query Match 40.0%; Score 963; DB 5; Length 409;
Best Local Similarity 52.2%; Pred. No. 1e-72;
Matches 177; Conservative 64; Mismatches 94; Indels 4; Gaps 3;

QY 104 EDDEEGHLCQSGDVARARVEIVDTLGEAGFAGKVCIDHGMGMMVAVKIVNNGRYRE 163
DB 54 QDDKDGHLIYSKDPFLINRTIYDTLGEFGKVR-VNDSISDTMALIKIKNSKYRE 112
QY 164 AARSEIQVLEHLNSTDPNSVRCQMLEWFDHGHGVCIVFELLGLSTYPIKENSPLPQ 223
DB 113 AAKLEVYVQLQKAEKQPEKGNWVIMHSGYFDYNGHICLLFDLMGSSIFDLKANHYPY 172
QY 224 IDHIRMAYQICQSIPLHNKLTHTDLPENILFVKSQDYVYKNSKMRDERTLKNTDI 283
DB 173 MEQTLHTWQCNVAVFLHDKLHTDLPENILFVDSRYTTLVD-KKPLRVLHSTHV 230
QY 284 KVVDPGSATYDDEHSTLVSTRHYRAPEVILALGMSQPCDWSIGCILLEYLGFVFT 343
DB 231 RLIDFGSATFDHHSITVSTRHYRAPEVILGWSQPCDWSIGCILLEYLGTGLFT 290
QY 344 HDSEKELAMERILGPDPQMIQTRKRRKFFHNQIDNDHSSAGRYVRRCKPLKEPVL 403
DB 291 HENREHLAMERVLGDIPLMARRT-KTFFINGRLDWNVTSADAAVYRDNCKPLRRSMS 349
QY 404 CHDEHEKLFDLVRMLEYDPTQITLDEALQHPFDLL 442
DB 350 CTDPHEVELFELIENMLPEPLARMKLPALQHRYPNRL 388

RESULT 15

035721 PRELIMINARY; PRT; 219 AA.
ID 035721
AC 035721:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cdc2/CDC28-like kinase 4 (Fragment).
GN CLK4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN (1)
RP SEQUENCE FROM N.A.
RA Watkins-Chow D.E., Douglas K.R., Buckwalter M.S., Probst F.J.,
RA Camper S.A.;
RT "Construction of a 3 Mb contig and transcript map of the central
RT region of mouse chromosome 11";
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U94846; AAB62179.1; -;
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
FT NON_TER 219
SQ SEQUENCE 219 AA; 25921 MW; CEAB6BBD2C2438C CRC64;

Query Match 34.8%; Score 839.5; DB 11; Length 219;
Best Local Similarity 85.6%; Pred. No. 1.1e-62;
Matches 160; Conservative 8; Mismatches 8; Indels 11; Gaps 2;

QY 8 SH-SVEEDTH-----PSHYLARSLSNERDYDRRYVDYRNDYCEGYVPRHYRD 56
DB 33 SHSSTQENRCKPHQFQKSDCHYLQARCLNERDYDRRYIDYRNDYCEGYVPRHYRD 92
QY 57 IESGYRIHCKSKSVSRSSPKRKRNRHCSHQSRKSRKRSRIEDDEGHLCQSG 116

DB 93 VESTYRIHCKSKSVSRSSPKRKRNRPCASHQSHSKSHRRKRSRIENDEGHLCQSG 152
QY 117 DVLRARVEIVDTLGEAGFAGKVCIDHGMGMMVAVKIVNNGRYREARARSETQVLEHLN 176
DB 153 DVLRARVEIVDTLGEAGFAGKVCIDHGMGMMVAVKIVNNGRYREARARSETQVLEHLN 212
QY 177 STDPNVS 183
DB 213 STDPNVS 219

Search completed: November 14, 2003, 15:18:41
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 03:23:10 ; Search time 7683.47 Seconds
(Without alignments)
19501.233 Million cell updates/sec

Title: US-10-028-946-1

Perfect score: 6165

Sequence: 1 atgttgagatcaaatatg.....gggaccagcttcagataaa 6165

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estha:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estcl:*
10: gb_estc2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pla:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssi:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	913.6	14.8	1011	13	BQ070955 AGENCOURT
2	822.6	13.3	879	13	BQ228524 AGENCOURT
3	814	13.2	956	13	BQ071141 AGENCOURT
4	810.2	13.1	830	13	BUI81633 AGENCOURT

5	750.6	12.2	849	12	BT253509	BT253509 602973370
6	732.8	11.9	920	11	BC342268	BC342268 BX342268
7	705.6	11.4	958	11	BC031156	BC031156 Mus muscu
8	698.2	11.3	843	12	BG912161	BG912161 602812833
9	629.2	10.2	652	10	BF905370	BF905370 IL3-MT026
10	628.8	10.2	640	13	BQ087302	BQ087302 NISC_K01
11	603.4	9.8	881	12	BG976452	BG976452 602846269
12	599	9.7	891	12	BT558919	BT558919 603240969
13	591.4	9.6	746	14	CA749290	CA749290 UT-M-FY0
14	588.6	9.5	681	9	AM05350	AM05350 QV3-DT004
15	581.6	9.4	968	13	BUI18506	BUI18506 603132657
16	575.2	9.3	1078	13	BUI139267	BUI139267 603132236
17	574.4	9.3	698	14	CD348416	CD348416 UT-M-FY0
18	566.4	9.2	1085	12	BM904785	BM904785 AGENCOURT
19	557.4	9.0	647	10	BB207065	BB207065 BB207065
20	538	8.7	566	2	BSM071348	BSM071348 Homo sapi
21	511.8	8.3	775	13	BUI05833	BUI05833 603005490
22	511	8.3	804	13	BUI236617	BUI236617 603411670
23	498.2	8.1	719	13	BUI39218	BUI39218 603515216
24	495.6	8.0	933	13	BUI232508	BUI232508 603408272
25	488	7.9	657	13	BQ081488	BQ081488 LB612149
26	481	7.8	820	13	BQ941506	BQ941506 AGENCOURT
27	451.4	7.3	459	13	BUI429033	BUI429033 UT-HF-BNO
28	450	7.3	600	12	BG808196	BG808196 2082-86 M
29	437.6	7.1	589	10	BF819093	BF819093 QV2-CI010
30	435.2	7.1	1075	13	BQ425970	BQ425970 AGENCOURT
31	431.2	7.0	497	10	BE909486	BE909486 601502838
32	424.2	6.9	662	13	BUI04008	BUI04008 603004106
33	423.8	6.9	427	12	BG944539	BG944539 ax51f11.x
34	417.2	6.8	749	13	BUI64866	BUI64866 603567832
35	416.6	6.8	678	13	BUI037714	BUI037714 604137885
36	414.8	6.7	875	13	BUI900029	BUI900029 AGENCOURT
37	412	6.7	617	13	BQ417810	BQ417810 IK5106.Y
38	399.8	6.5	1013	13	BUI08983	BUI08983 603110947
39	396	6.4	703	14	CB512917	CB512917 ssal19b53
40	394.8	6.4	936	12	BT758420	BT758420 603026948
41	387.8	6.3	720	14	CA380902	CA380902 660300 NC
42	380.4	6.2	463	14	CB735389	CB735389 AGENCOURT
43	376	6.1	637	14	BY713457	BY713457 BY713457
44	374.4	6.1	522	14	T87377	T87377 Yd83e05.r1
45	373.4	6.1	473	10	BB728630	BB728630 BB728630

ALIGNMENTS

RESULT 1
LOCUS BQ070955
DEFINITION AGENCOURT 6855647 NIH_MGC_47 Homo sapiens CDNA clone IMAGE:5923441
5' mRNA sequence.
ACCESSION BQ070955
VERSION BQ070955.1 GI:19900001
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1011)
NIH-MGC <http://nigc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1CM2091 row: p column: 02
High quality sequence step: 634.

FEATURES

Location/Qualifiers
 1..1011
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5923441"
 /tissue_type="neuroblastoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_47"
 /note="Organ: Brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library." 4 others

BASE COUNT 296 a 248 c 307 g 156 t

ORIGIN

Query Match 14.8%; Score 913.6; DB 13; Length 1011;
 Best Local Similarity 97.8%; Pred. No. 2e-206; Indels 3; Gaps 2;
 Matches 945; Conservative 0; Mismatches 18;

2701 CGGAGGTCAGTCTAGAGCAGAGCAGAACTGAGCTCAAGCCAGCTCAGAG 2760
 1 CGGAGGTCAGTCTAGAGCAGAGCAGAACTGAGCTCAAGCCAGCTCAGAG 60

2761 CTACAGCTCTCCCTGAGAGCGCGAGTCAAGTTCAGCCCTGAGGCTGACGGGG 2820
 61 CTACAGCTCTCCCTGAGAGCGCGAGTCAAGTTCAGCCCTGAGGCTGACGGGG 120

2821 GCCCTGAGAGCAGCTTCCGCGAGGAGACAGAGCTGAAAGAGCAGCAGAGAGCT 2880
 121 GCCCTGAGAGCAGCTTCCGCGAGGAGACAGAGCTGAAAGAGCAGCAGAGAGCT 180

2881 GAAGAGAGATCCAGGACTCAGGCACTAGAGATGAATCCAGGCAATTGATGCT 2940
 181 GAAGAGAGATCCAGGACTCAGGCACTAGAGATGAATCCAGGCAATTGATGCT 240

2941 CTTCGTACAGCTGTACTGTATCAGAGACTGAGAGAGAGTAAACAGCTGACCGAG 3000
 241 CTTCGTACAGCTGTACTGTATCAGAGACTGAGAGAGAGTAAACAGCTGACCGAG 300

3001 GACACGCTGAATCAACAAACAACTTCTACTGTCCAAACAACTGATGAGGCTTCT 3060
 301 GACACGCTGAATCAACAAACAACTTCTACTGTCCAAACAACTGATGAGGCTTCT 360

3061 GGCGCCAGAGAGAGTGTACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3120
 361 GGCGCCAGAGAGAGTGTACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 420

3121 ACCGAAAGAGAGATGAGCTTACACAGCAGAAACAGATGAGAGCTTGAAGACAG 3180
 421 ACCGAAAGAGAGATGAGCTTACACAGCAGAAACAGATGAGAGCTTGAAGACAG 480

3181 TGACACATGCTGAGAGAAAGAGTCAAGATTGAGAGCCCTAAACATGAGCTGTAGAA 3240
 481 TGACACATGCTGAGAGAAAGAGTCAAGATTGAGAGCCCTAAACATGAGCTGTAGAA 540

3241 AAAGAGCGAGTGGAGAGCTTGAAGAGAGCTCTGAGTGAAGAAATCCAGATTGAG 3300
 541 AAAGAGCGAGTGGAGAGCTTGAAGAGAGCTCTGAGTGAAGAAATCCAGATTGAG 600

3301 TGTGCGGTTTGAAGAGTCTGAGAGAGATGCTGAGACCGAAGACAGGCGGAGAGCC 3360
 601 TGTGCGGTTTGAAGAGTCTGAGAGAGATGCTGAGACCGAAGACAGGCGGAGAGCC 660

3361 GATCAGCGAGTCAAGAGTCTCGCAGAGTGTGAGAGCTGAGTGAAGAGCAGAGGCT 3420
 661 GATCAGCGAGTCAAGAGTCTCGCAGAGTGTGAGAGCTGAGTGAAGAGCAGAGGCT 720

3421 GAGATTCTCGCTTCGAGAGGCTCTGAAAGACAGAAAGTGAAGCCGAGAGCTTCT 3480
 721 GAGATTCTCGCTTCGAGAGGCTCTGAAAGACAGAAAGTGAAGCCGAGAGCTTCT 780

3481 GACAGGTCATGACCTCGAGAGAGAGCAGTATGCTTGAATGAATGCC- GAAGCTT 3539
 781 GACAGGTCATGACCTCGAGAGAGAGCAGTATGCTTGAATGAATGCCGGAAGCTT 840

3540 ACAGAGAGTGGAGAGCTGAGAGAGTCTGAAACAGAGGCTTCTGAGAGAGAGCCAA 3599
 841 ACAGAGAGTGGAGAGCTGAGAGAGTCTGAAACAGAGGCTTCTGAGAGAGAGCCAA 900

3600 ATTACAGAGAGATGAGAGCTGAGAGAGTCTGAAATGACATTTT-CCGTGACTCAGAGCTG 3657
 901 ATTACAGAGAGATGAGAGCTGAGAGAGTCTGAAATGACATTTTCCGTGACTCAGAGCTG 960

3658 CAGAGA 3663
 961 CCAGAA 966

RESULT 2
 BQ228524
 LOCUS
 DEFINITION
 AGENCOURT 7600932 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5061485
 5', mRNA sequence.
 ACCESSION
 BQ228524
 BQ228524.1 GI:20409924
 EST.
 SOURCE
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 879)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC/DCTD/DP
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LHAM3331 row: 9 column: 22
 High quality sequence stop: 636.

FEATURES
 source
 Location/Qualifiers
 1..879
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5061485"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_72"
 /note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies." 1 others

BASE COUNT 221 a 258 c 239 g 160 t 1 others

ORIGIN

Query Match 13.3%; Score 822.6; DB 13; Length 879;
 Best Local Similarity 98.0%; Pred. No. 9.1e-185;
 Matches 853; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

3949 CGGAGAGAGTGGCCACCGCAAGCAACGAGCAGCAGCAGCAGCAGCAGCAGCAGC 4008
 1 CGGAGAGAGTGGCCACCGCAAGCAACGAGCAGCAGCAGCAGCAGCAGCAGCAGC 60

4009 GCGAGGAGAGATGCGCATGTCGCGCATGTCGCGCATGTCGCGAGAGACAGCCAGTGC 4068

Db	61	GGGAGGACGAGATGCGCATGTCGGCATTCGGTGGCCAGACACACAGCCAGTCC	120
QY	4069	ATGAGCCTGCTGCGCCCGCATCCAGCCGACGAAAGAGTCTTCACTCCAGAGATT	41228
Db	121	ATGAGCCTGCTGCGCCCGCATCCAGCCGACGAAAGAGTCTTCACTCCAGAGATT	180
QY	4129	AGTGGGCGCTTAAGAAACGATGACACCAATATCTCACCCGTTCAAGTAGACG	4188
Db	161	AGTCGGCGCTTTAAGAAACGATGACACCAATATCTCACCCGTTCAAGTAGACG	240
QY	4189	AACATGCGAGCCACAAAGTGTGCTGTGTCTGATACCGTGCATTTTGAACGACGCA	4248
Db	241	AACATGCGAGCCACAAAGTGTGCTGTGTCTGATACCGTGCATTTTGAACGACGCA	300
QY	4249	TTCCAAATGCTCTGAAATGTTCAGGTATGTGTCAACCCCAAGTGTCCAGTGTTCGACCC	4308
Db	301	TTCCAAATGCTCTGAAATGTTCAGGTATGTGTCAACCCCAAGTGTTCAGTGTTCGACCC	360
QY	4309	ACCTGCGGCTTGCCTGCTGTAATATCCACACACTTCACCGAGGCTTTCGCGGTACAA	4368
Db	361	ACCTGCGGCTTGCCTGCTGTAATATCCACACACTTCACCGAGGCTTTCGCGGTACAA	420
QY	4369	ATGAATCTCCAGAGTCTTCACACACAGAGCCACAGAGCTTGCACCTTGAGAGGTGG	4428
Db	421	ATGAATCTCCAGAGTCTTCACACACAGAGCCACAGAGCTTGCACCTTGAGAGGTGG	480
QY	4429	ATGAAGTGGCCAGGAATTAACAAACGAGAGACGAAAGCTGGGACAGGAATGATGTC	4488
Db	481	ATGAAGTGGCCAGGAATTAACAAACGAGAGACGAAAGCTGGGACAGGAATGATGTC	540
QY	4489	CTGAGGAGATCAAAAGTCTCATTTATGACAATGAAGCCAGAGAGCTGACAGAGCGG	4548
Db	541	CTGAGGAGATCAAAAGTCTCATTTATGACAATGAAGCCAGAGAGCTGACAGAGCGG	600
QY	4549	GTCGAGAGATTGAGCTGTGCTTCGACGCGGATGATATCATTCATGAGGCGTGTGCT	4608
Db	601	GTCGAGAGATTGAGCTGTGCTTCGACGCGGATGATATCATTCATGAGGCGTGTGCT	660
QY	4609	GCTTCGAACTCGCAAAATPACAGCCCAAGACAGATGTCCCATCATCATCTGAAATGAACT	4668
Db	661	GCTTCGAACTCGCAAAATPACAGCCCAAGACAGATGTCCCATCATCATCTGAAATGAACT	720
QY	4669	CACCGGCACACCACTGTCT-GGCCGGGAGAACCTCTTAATCTTGCTGCTCCAGCTTCC	4727
Db	721	CNACCGCACACCACTGTGTGGGCCCGGAGAACTCTTAATCTTGCTGCTCCAGCTTCC	780
QY	4728	TGACAAACAGCGCTGGTCAACCGCTTGAATCAATGTGCGAGGT-GGAGAGTTTCTA	4786
Db	781	TGACAAACAGCGCTGGTCAACCGCTTGAATCAATGTGCGAGGTGGAGAGATTCTA	840
QY	4787	GCGAAAGACGAGACCTGATGCTAAACTGC	4816
Db	841	GCGAAAGACGAGACCTGATGCTGCGCG	870
RESULT 3			
LOCUS	BO071141	956 bp	mRNA
DEFINITION	AGNCOURT_6853098 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5927502	linear	EST 02-APR-2002
ACCESSION	BO071141		
VERSION	BO071141.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 956)		
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Strausberg, Ph.D.		

		Email: cga@bs-remail.nih.gov
		Tissue Procurement: ATCC
		cDNA Library Preparation: Rubin Laboratory
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
		DNA Sequencing by: Agencourt Bioscience Corporation
		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
		http://image.lnl.gov
		Plate: LNCM2102 row: i column: 07
		High quality sequence stop: 650.
FEATURES		Location/Qualifiers
source		1..956
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:5927502"
		/tissue_type="neuroblastoma, cell line"
		/lab_host="DH10B (phage-resistant)"
		/clone_lib="NIH MGC 47"
		/note="Organ: brain; Vector: pOT8; Site 1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >50bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT	283 a	238 c
ORIGIN		292 g 143 t
Query Match	13.2%	Score 814; DB 13; Length 956;
Best Local Similarity	97.0%;	Pred. No. 1,18-182;
Matches 873; Conservative	0; Mismatches 20; Indels 7; Gaps 4;	
QY	2701	CGGAGGTCTAAGTCATGACGACGAGAGCAGAAACTGAGCTCAAGCCGACCTCACAGAG 2766
DB	1	CGGAGGTCTAAGTCATGACGACGAGAGCAGAAACTGAGCTCAAGCCGACCTCACAGAG 60
QY	2761	CTACAGCTCTCCCTGCGAGAGCGGAGCTCATGACAGCCCTGCGAGGCTGACAGCGGCG 2820
DB	61	CTACAGCTCTCCCTGCGAGAGCGGAGCTCATGACAGCCCTGCGAGGCTGACAGCGGCG 120
QY	2821	GCCCTGAGAGCCAGCTTGCGCAGGCGGAACAAGAGCTGGAAGAACCACAGCAGAAAGCT 2880
DB	121	GCCCTGAGAGCCAGCTTGCGCAGGCGGAACAAGAGCTGGAAGAACCACAGCAGAAAGCT 180
QY	2881	GAAAGAGAGATCCAGGGACCTCAGGCGCATGAGATGAATAAACGCCGAATTGATGCT 2940
DB	181	GAAAGAGAGATCCAGGGACCTCAGGCGCATGAGATGAATAAACGCCGAATTGATGCT 240
QY	2941	CTTGTAACACTCTGTAAGTATCAACAGACCTTGAGAGAGCAGCTAAACAGACTGACCGAG 3000
DB	241	CTTGTAACACTCTGTAAGTATCAACAGACCTTGAGAGAGCAGCTAAACAGACTGACCGAG 300
QY	3001	GACAAAGCTGAATCTCAACACCAAACCTTTCTACTTGTCCAAACAACTCGATGAGAGCTTCT 3060
DB	301	GACAAAGCTGAATCTCAACACCAAACCTTTCTACTTGTCCAAACAACTCGATGAGAGCTTCT 360
QY	3061	GAGCCCAACGACGAGATTTGTAACAATCGAAGTAAATGAGCAATCTCCGCGGGAAGATC 3120
DB	361	GAGCCCAACGACGAGATTTGTAACAATCGAAGTAAATGAGCAATCTCCGCGGGAAGATC 420
QY	3121	ACGGAACGAGATGACAGCTTACAGGCGAAGCAAAAGATGAGAGGCTCTGGAAGCAGCAG 3180
DB	421	ACGGAACGAGATGACAGCTTACAGGCGAAGCAAAAGATGAGAGGCTCTGGAAGCAGCAG 480
QY	3181	TGCACCATGCTGGAGGAACAGGTCATGAGATTGGAGGCGCTTAACCATGAGACTCTGGA 3240
DB	481	TGCACCATGCTGGAGGAACAGGTCATGAGATTGGAGGCGCTTAACCATGAGACTCTGGA 540
QY	3241	AAGAGCGGACGATGAGAGGCTCTGAGAGAGCGTCTGAGGTATGAGAAATCCAGATTGAG 3300

Db 541 AAAGAGCGGAGGAGGCTGAGAGAGCTCTGGGTGATGAGAAATCCAGTTTGAG 600
QY 3301 TGTGGGTTTCGAGAGCTGCGAGAGATGCTTGGACACCGAGAAACAGAGCGAGGCGAGCC 3360
Db 601 TGTGGGTTTCGAGAGCTGCGAGAGATGCTTGGACACCGAGAAACAGAGCGAGGCGAGCC 660
QY 3361 GATCAGCGGATCAGCGAGCTGCGAGAGTGTGTGAGCTGCGAGAGTGTGAGAGCAACAAGCT 3420
Db 661 GATCAGCGGATCAGCGAGCTGCGAGAGTGTGTGAGCTGCGAGAGTGTGAGAGCAACAAGCT 720
QY 3421 GAGATTCTGCTGTGCGAGAGCTCTCAAAAGCAGAACTGAGAGCGGAGAGCTCTCT 3480
Db 721 GAGATTCTGCTGTGCGAGAGCTCTCAAAAGCAGAACTGAGAGCGGAGAGCTCTCT 780
QY 3481 GACAAAGCTCAATGACCT-GGAGAGAGAGATGCTTGTGAA--TGATGCGCCGAGC 3537
Db 781 GACAAAGCTCAATGACCTGGAGAGAGAGATGCTTGTGAAATGATGCGCCGAGC 840
QY 3538 TTACAGCAGAA--GCTGAGACTGAAC--GAGAGCTCAAAACAGAGCTTGTGAGAGCA 3593
Db 841 TTACAGCAGAAAGCTGGAGACTGAACCGAGAGCTCAAAACAGAGCTTGTGAGAGCA 900

RESULT 4
BUI1633 830 bp mRNA linear EST 04-SEP-2002
LOCUS AGENCOURT 7906225 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6140538
DEFINITION 5', mRNA sequence.
ACCESSION BUI1633
VERSION BUI1633.1 GI:22695617
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 830)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNI at:
http://image.llni.gov
Plate: LAM13459 row: e column: 19
High quality sequence stop: 652.
location/Qualifiers
1..830
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6140538"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-Sport6; Site: 1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 1.75 kb. Library constructed by Life
Technologies."

BASE COUNT 259 a 202 c 239 g 129 t 1 others
ORIGIN

Query Match 13.1%; Score 810.2; DB 13; Length 830;
Best Local Similarity 99.4%; Pred. No. 8e-182;
Matches 823; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 2435 TGGATTCCAAAGATCGATCCCTGGAAGAGAGATGTTGAACTGCTGAGCCATATAAC 2494
Db 1 TGGATTCCAAAGATCGATCCCTGGAAGAGAGATGTTGAACTGCTGAGCCATATAAC 60

QY 2495 TTGCAAGTAATACAGTCTTTTAAACCAAGAAATGAAAGGCTCCAGAGATGATTT 2554
Db 61 TTGCAAGTAATACAGTCTTTTAAACCAAGAAATGAAAGGCTCCAGAGATGATTT 120
QY 2555 CTGAAGCTCAAGCAACAGAAATTTTAACTGAGACACAGAGCTGGAGAGTTGAGAGCCAG 2614
Db 121 CTGAAGCTCAAGCAACAGAAATTTTAACTGAGACACAGAGCTGGAGAGTTGAGAGCCAG 180
QY 2615 ACCGAAATCTGAGAGACAGCTGAGAGATTCAGCCACCAAGACCAACAGTGAACAATC 2674
Db 181 ACCGAAATCTGAGAGACAGCTGAGAGATTCAGCCACCAAGACCAACAGTGAACAATC 240
QY 2675 GGCTGCTGGAATCGAGACAGATTCGAGAGAGTCAAGTCAAGACCAAGACCAAGATC 2734
Db 241 GGCTGCTGGAATCGAGACAGATTCGAGAGAGTCAAGTCAAGACCAAGACCAAGATC 300
QY 2735 TGGAGCTCAAGCCGACAGCTCAGAGACTACAGCTCTCCCTGACAGAGCGGAGTCAAGT 2794
Db 301 TGGAGCTCAAGCCGACAGCTCAGAGACTACAGCTCTCCCTGACAGAGCGGAGTCAAGT 360
QY 2795 TGAAGCTCTGAGAGCTGACAGCGGCGGCTTGGAGAGCCAGCTTCCGACAGCGAAGAC 2854
Db 361 TGAAGCTCTGAGAGCTGACAGCGGCGGCTTGGAGAGCCAGCTTCCGACAGCGAAGAC 420
QY 2855 AGCTGGAAGAGACCAAGAGAGTGAAGAGAGATCCAGGACTCAGGCAATAGAG 2914
Db 421 AGCTGGAAGAGACCAAGAGAGTGAAGAGAGATCCAGGACTCAGGCAATAGAG 480
QY 2915 ATGAAATCCAGGCAATTTGATGCTCTTCTGTAACAGCTGTACTGTAATCAAGACCTGG 2974
Db 481 ATGAAATCCAGGCAATTTGATGCTCTTCTGTAACAGCTGTACTGTAATCAAGACCTGG 540
QY 2975 AGGAGAGCTCAACCAAGAGAGTGAAGAGAGATCCAGGACTCAGGCAATAGAG 3034
Db 541 AGGAGAGCTCAACCAAGAGAGTGAAGAGAGATCCAGGACTCAGGCAATAGAG 600
QY 3035 TGTCCAAACCACTGATGAGAGCTTGTGGCGCCAAAGAGAGATTTGAACTCGGAAGTG 3094
Db 601 TGTCCAAACCACTGATGAGAGCTTGTGGCGCCAAAGAGAGATTTGAACTCGGAAGTG 660
QY 3095 AAGTGGACATCTCCGCGGAGAGATCAGAGAAACAGAGATGAGCTTACAGCAGAGAG 3154
Db 661 AAGTGGACATCTCCGCGGAGAGATCAGAGAAACAGAGATGAGCTTACAGCAGAGAG 720
QY 3155 AAACGATGAGAGCTCTGAGAGACCAAGCTGACCATGCTGAGAGAAACAGTCAATGATT 3213
Db 721 AAACGATGAGAGCTCTGAGAGACCAAGCTGACCATGCTGAGAGAAACAGTCAATGATT 780
QY 3214 GAGGCTTAAACGATGAGCTGCTGAGAGAAAGAGCGGCAAGTGGAGGCGC 3261
Db 781 GAGGCTTAAACGATGAGCTGCTGAGAGAAAGAGCGGCAAGTGGAGGCGC 828

RESULT 5
B1253509 849 bp mRNA linear EST 17-JUL-2001
LOCUS B1253509
DEFINITION 60297370F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112737 5',
mRNA sequence.
ACCESSION B1253509
VERSION B1253509.1 GI:14805003
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 849)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMT at: <http://www.imgc.org/>

FEATURES	Location/Qualifiers
source	1. ,849

BASE COUNT	218 a	230 c	211 g	190 t
ORIGIN				

Query Match	12.2%;	Score 750.6;	DB 12;	Length 849;
Best Local Similarity	98.5%;	Pred. No. 1.3e-167;		
Matches 789; Conservative	0;	Mismatches 9;	Indels 3;	Gaps 3;

QY	4812	ACTGCTTGAAACCTCCGCTGAAACTGGAAAGTATACCGCTAGACATGAACCTGAC	487
Db	3	AAGCGTCGGAAACCTCCGCTGAAACTGGAAAGTATACCGCTAGACATGAACCTGAC	62
QY	4872	GCTGCCCTTCACTGACACAGTGTGTGTGTGGGACCCGAGAAAGGCGCTTACGGCCCTGA	493
Db	63	GCTGCCCTTCACTGACACAGTGTGTGTGTGGGACCCGAGAAAGGCGCTTACGGCCCTGA	122
QY	4932	TGCTTTGAAAACTCCCTPAACCCATGTCCCAAGAAATTGGAGCAGCTTCCAAATTATAT	499
Db	123	TGCTTTGAAAACTCCCTPAACCCATGTCCCAAGAAATTGGAGCAGCTTCCAAATTATAT	182
QY	4992	TATCAAGGACCTTGGAGAGCTTCTATGTATACAGAGAAAGAGCGGCACTGTCTTGT	505
Db	183	TATCAAGGACCTTGGAGAGCTTCTATGTATACAGAGAAAGAGCGGCACTGTGTCTTGT	242
QY	5052	GAGCGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCGACCTGCTGGCCAGCCGACAT	511
Db	243	GAGCGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCGACCTGCTGGCCAGCCGACAT	302
QY	5112	CTACCCCAACATTTTGAAGCTGTGAAGGCGTGCACATGTTTGGGGGACAGGAAATGA	517
Db	303	CTACCCCAACATTTTGAAGCTGTGAAGGCGTGCACATGTTTGGGGGACAGGAAATGA	362
QY	5172	GAACGGGCTGTGCATCTGTGACGCCATGCCGAAAGTGTGTATCTTCGCTACAAAGA	523
Db	363	GAACGGGCTGTGCATCTGTGACGCCATGCCGAAAGTGTGTATCTTCGCTACAAAGA	422
QY	5232	AAACCTCAGCAAAATCTGCATCCGGAAAGAGATAGAGCTCAGAGCCCTGACAGCTAT	529
Db	423	AAACCTCAGCAAAATCTGCATCCGGAAAGAGATAGAGCTCAGAGCCCTGACAGCTAT	482
QY	5292	CCACTTCAACCAATTACAGTATCTCATTGGAACCAATAAATTCTACGAATGACATGAA	535
Db	483	CCACTTCAACCAATTACAGTATCTCATTGGAACCAATAAATTCTACGAATGACATGAA	542
QY	5352	GGAGTACACGCTCGAGGAATTCCTGTATAGAAATAGCAATTCCTTGGAGCCGCGTGT	541
Db	543	GGAGTACACGCTCGAGGAATTCCTGTATAGAAATAGCAATTCCTTGGAGCCGCGTGT	602
QY	5412	TGCGGCTCTTTCACACAGCTTCCCTGTCTCAATCTGTGACAGTGAACAGCGCAGGACG	547
Db	603	TGCGGCTCTTTCACACAGCTTCCCTGTCTCAATCTGTGACAGTGAACAGCGCAGGACG	662

Accession	Sequence	Position
QY	5412 AGAGAGACTGCTGTGTTCCAGCAATTTGAGTGTGTGTGATTCCTACGGAAGCG	5531
Db	663 AGAGAGACTGCTGTGTTCCAGCAATTTGAGTGTGTGTGATTCCTACGGAAGCG	722
QY	5532 T-AGCCGACAGCAGATCTCAAGTGGAGTGCTTACCTTTGGCCTTGGCTTACAGAGAC	5599
Db	723 TAAGCCGACAGCAGATCTCAAGTGGAGTGCTTACCTTTGG-CTTTGCCACAGAGA-	780
QY	5581 CCTATCTGTTGTGACCCACT 5611	
Db	781 CCTATCTGTTGGAGCCATT 801	

[illegible]

FEATURES	Location/Qualifiers
source	1. .920

BASE COUNT	216 a	298 c	249 g	157 t
ORIGIN				

Query Match	11.94;	Score 732.8;	DB 13;	length 920;
Best Local Similarity	99.64;	Pred. No. 2.4e-163;		
Matches 745; Conservative	0;	Mismatches 2;	Indels 1;	Gaps 1;

[illegible]

QY 5598 GTTTGTACCCACTTCACTCACTCGAAGTAATTGATTCAGACACGCTCTCTACAGAG 5657
 Db 246 GTTGTACCCACTTCACTCACTCGAAGTAATTGATTCAGACACGCTCTCTACAGAG 305
 QY 5658 GACCCCTGCGGAGGAGTACCTGAGCATCCGGAACCCGCGTACCTGAGGCTCTGCAATTC 5717
 Db 306 GACCCCTGCGGAGGAGTACCTGAGCATCCGGAACCCGCGTACCTGAGGCTCTGCAATTC 365
 QY 5718 CTGAGGAGGATTTACTTGGCGTCTCTATCCAGGATTAATTGAGGTCATTTGCTGCAA 5777
 Db 366 CTGAGGAGGATTTACTTGGCGTCTCTATCCAGGATTAATTGAGGTCATTTGCTGCAA 425
 QY 5778 GGGAAACCTGCTGAAGAGAGTCCGCGCACTGAACACACCGGAGCCGCTCCACCTCCGAG 5837
 Db 426 GGGAAACCTGCTGAAGAGAGTCCGCGCACTGAACACACCGGAGCCGCTCCACCTCCGAG 485
 QY 5838 CAGCCCAAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 5897
 Db 486 CAGCCCAAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 545
 QY 5898 CAGCCCAAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 5957
 Db 546 CAGCCCAAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 604
 QY 5958 CCGGAGAGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 6017
 Db 605 CCGGAGAGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 664
 QY 6018 GAATCTCCGCGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 6077
 Db 665 GAATCTCCGCGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 724
 QY 6078 AGACAGAGAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 6137
 Db 725 AGACAGAGAGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 784
 QY 6138 CAAGGTCGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 6165
 Db 785 CAAGGTCGGAGAGAGAGGAGCCCAAGGATCAACAGAGACATCAACAGAGGCGTGGCTTC 812

RESULT 7
 BC031156
 LOCUS BC031156 958 bp mRNA linear HTC 04-MAR-2003
 DEFINITION Mus musculus, similar to citron, clone IMAGE:4976752, mRNA.
 ACCESSION BC031156 GI:21411076
 VERSION BC031156.1
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 958)
 Strausberg, R.
 Direct Submission
 Submitted (03-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdeparil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 59 Row: 3 Column: 6
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis
 This clone has the following problem: retained intron.

FEATURES
 source location/qualifiers
 1..958
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4976752"
 /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
 ductal carcinoma, 5 month old virgin mouse."
 /clone_id="NCI_GAP_Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

BASE COUNT 258 a 288 c 242 g 170 t
 ORIGIN

Query Match 11.4%; Score 705.6; DB 11; Length 958;
 Best Local Similarity 89.1%; Pred. No. 7.7e-157;
 Matches 775; Conservative 0; Mismatches 89; Indels 6; Gaps 1;

QY 5302 AATTACAGTATCCCTCATTTGAACCAATTAATTACAGAAATCGACATGAAGCAGTACAG 5361
 Db 40 AAGAAAGCATCTCTCATTTGAGACCAAAATTTCTATGATCGACATGAAGCAGTACAG 99
 QY 5362 CTGAGAGATTTCTGATTAAGATGACCAATTTCTTGGCACTGCTGTTTCCGCTCT 5421
 Db 100 CTGATGATTTCTGAGACAGACACATTTCTTGGCACTGCTGTTTCCGCTCTG 159
 QY 5422 TCCAAACAGCTTCCCTGCTCATGCTGAGGAGTGAACAGCGAGCGAGAGAGATAC 5481
 Db 160 TCCAAACAGCTTCCCTGCTCATGCTGAGGAGTGAACAGCGAGCGAGAGAGATAC 219
 QY 5482 TTGCTGTGTTTCAAGAAATTTGAGTGTCTGAGATTTCTTAACGAGAGCGGACACA 5541
 Db 220 CTGCTGTCTTCCAGAAATTTGAGTGTCTGAGATTTCTTAACGAGAGCGGACACA 279
 QY 5542 GACGATCTCAAGTGAAGTGTCTTACCTTTGGCTTGGCTTACAGAAACCTTCTGTT 5601
 Db 280 GATGATCTTAAAGTGAAGTGTCTTACCTTTGGCTTGGCTTACAGAAACCTTCTGTT 339
 QY 5602 GTGACCCACTTCACTCACTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAGAC 5661
 Db 340 GTGACCTCACTTCACTCTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAGAC 399
 QY 5662 CTTGCGGAGGATCTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAGAC 5721
 Db 400 CTTGCGGAGGATCTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAGAC 459
 QY 5722 GAGAGGATTTCTGAGGATCTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAG 5781
 Db 460 GAGAGGATTTCTGAGGATCTGAGATTAATTGAGATCCAGGACGCTCTCAGCAGGAG 519
 QY 5782 AAAGCTGTGAAGAGTCCGAGCATGAACACCCGAGGAGCGTCACTCTCCGAGAGC 5841
 Db 520 AAAGCTGTGAAGAGTCCGAGCATGAACACCCGAGGAGCGTCACTCTCCGAGAGC 579
 QY 5842 CCGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5901
 Db 580 CCGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639
 QY 5902 CCGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5959
 Db 640 CCGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 699
 QY 5960 ----CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6015
 Db 700 GAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 759

QY	6016	GAGAAAGTCCCCCGGCGCGATATCTCAGCAACGCGAAGAGCGATCCCCCGCAGAGGCTGTTT	6073
Dd	760	GAGAAAGTCCCCCGGCGCGAATGCTCAGCACTTGGAGAGAGCGGATCCCCGAGAGACTGTTT	819
QY	6076	GAAAGCACACAGCAGGGGCGCGGCTGCTCGGAGACGCTGAGAGACCCCGCTGTCACAGGTG	6135
Dd	820	GAAAGCACACAGCAGGGGCGCGGCTGCTCGAAGAGACGTGAGAGACCCCACTGTCCAGGTT	879
QY	6136	AACAAGGCTCTGGAGACCAAGTCTTCAATATTA	6165
Dd	880	AACAAGGCTCTGGAGACCAAGTCTTCAATATTA	909

RESULT 8
LOCUS BG912161
DEFINITION BG912161 843 bp mRNA linear EST_05-JUN-2001
602812833P1 MCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4944657
5'', mRNA sequence.
ACCESSION BG912161
VERSION BG912161.1 GI:14292637
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 843)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Tissue Procurement: David N. Louis, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
 Plate: LIML10890 row: a column: 10
 High quality sequence stop: 778.

FEATURES	
source	Location/Qualifiers
	1..843
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone_image="4944657"
	/issue_type="anaplastic oligodendroglioma with 1p/19q loss"
	/lab_host="DH10B (T1 phage-resistant)"
	/clone_lib="NCI CGAP Brn67"
	/note="Organ: brain; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT	218 a 227 c 214 g 164 t
ORIGIN	

	Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
	11.3%; Score 698.2; DB 12; Length 843;	93.3%; Pred. 0.4, 2.e-155;	0;	48;	8;	5;	
QY	4669	CACCCGACACCACTCTGTGGCCCGGGGAAACCTCTACTGTTGTAAGTCCGAGCTTCCCT	4728				
Db	2	CACCGCTGCACCACTCTGTGGCCCGGGGAAACCTCTACTGTTGTAAGTCCGAGCTTCCCT	61				
QY	4729	GACAAAGAGCGCTGGGTCCACCGCCCTTGAATACAGTTGTGCGAGGTGGAGAGTTTCTAGG	4788				
Db	62	GACAAAGAGCGCTGGGTCCACCGCCCTTGAATACAGTTGTGCGAGGTGGAGAGTTTCTAGG	121				
QY	4789	GAAAGACGAGAGCTGATGTAAACTGCTTGAAGTCCCTGCGGAACTGGAAGGTCAT	4848				
Db	122	GAAAGACGAGAGCTGATGTAAACTGCTTGAAGTCCCTGCGGAACTGGAAGGTCAT	181				

QY	4849	GACCGTATGACATGAATGACAGCTGACCGCTTGATGATACCAAGGGTGGTGGTGGACACC	4908
Dp	182	GACCGTATGACATGAATGACAGCTGACCGCTTGATGATACCAAGGGTGGTGGTGGACACC	241
QY	4909	GAGGAAGGCGCTTACGCGCCCTGAATGTCTTGAAAACTCCCTAACCCATGTGCCAGAAATT	4968
Dp	242	GAGGAAGGCGCTTACGCGCCCTGAATGTCTTGAAAACTCCCTAACCCATGTGCCAGAAATT	301
QY	4969	GGAGCAGTCTTCCAAATTTATATTAATGAAGACCTGGAGAAGCAGTCAATGATGAGAGA	5028
Dp	302	GGAGCAGTCTTCCAAATTTATATTAATGAAGACCTGGAGAAGCAGTCAATGATGAGAGA	361
QY	5029	GAAAGAGCGGACACTGTGTCTTGATGACGTGAAGAAATGAAAACAGTCCCTGCGCCAGTCC	5088
Dp	362	GAAAGAGCGGACACTGTGTCTTGATGACGTGAAGAAATGAAAACAGTCCCTGCGCCAGTCC	421
QY	5089	CACCTGCTGCGCCACGCGCCGACATCTCACCCCAACTTTTGAAGCTGTCAAGAGCGCTGCAC	5148
Dp	422	CACCTGCTGCGCCACGCGCCGACATCTCACCCCAACTTTTGAAGCTGTCAAGAGCGCTGCAC	481
QY	5149	TTGTGTTGAGGACAGCAAGATTGAGAAACGGGCTGTGACTGTGTGAGGCATGCCAGCAAA	5208
Dp	482	TTGTGTTGAGGACAGCAAGATTGAGAAACGGGCTGTGACTGTGTGAGGCATGCCAGCAAA	541
QY	5209	GTGCTCAATTCGCGCTACACGAAACCTGACGAAATACATGCAATCCGGAATGACATATGAG	5268
Dp	542	GTGCTCAATTCGCGCTACACGAAACCTGACGAAATACATGCAATCCGGAATGACATATGAG	601
QY	5269	ACCTGAGAGCCCTGACAGCTGTATTCACCTTCAACCAATTACAGTATCTTATT - GGAACCAA	5327
Dp	602	ACCTGAGAGCCCTGACAGCTGTATTCACCTTCAACCAATTACAGTATCTTATT - GGAACCAA	661
QY	5328	TAAATTTATAG - AAATGAGATGAA - GCAGTACACGCTGACAGGAAT - - CCGTAGATTAGA	5383
Dp	662	TAAATTTATAGAAATGAGATGAGAGCAGTACACGCTGACAGGAATTTCTGATTAACGA	721
QY	5384	ATGACCAATTCCTTGGACCTGCTGTG - - TTTGCGGCTCTTCCACAGACTTCCCTGTCT	5440
Dp	722	ATGACCAATTCCTTGGACCTGAGCTGAGTTTAAAGCGGCTCTTCCAAAAGTTCTGTCTCTC	781
QY	5441	CAATTCGACGAGTGAACAGCGACAGGCGACGAGGAGAGAGTACTTGTCTGTGTTCCAGCAAT	5500
Dp	782	AATCGTGGCAGGTGAACAGGCGACGCGACGAGGAGAGAGTATTCGTGTGTTCCCGGACT	841
QY	5501	T 5501	
Dp	842	T 842	

RESULT	9
LOCUS	BF905370/c
DEFINITION	652 bp mRNA linear EST 18-JAN-2001 IL3-MT0267-261200-410-H07 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BF905370
VERSION	BF905370.1 GI:12296829
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 652) Nagai,N.E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Borda,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsushima,A., Baia,G.S., Simpson,D.H., Brumsteih,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare, 'M.J.', Soares,F., Brentani,R.R., Reis,I.P., de Souza,S.J. and Simpson,A.U.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc.Natl. Acad. Sci. U.S.A. 97 (??), 3491-3496 (2000)
MEDLINE	20202663
PUBMED	10737800
COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=1l3&t2=1l3-MT0267-
261200-410-H07&t3=2000-12-26&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 624.
Location/Qualifiers
1. .652
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0267"
/note="Organ: marrow; Vector: puc18; Site 1: Sma1; Site 2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 94 a 189 c 163 g 206 t
ORIGIN

Query Match 10.2%; Score 629.2; DB 10; Length 652;
Best Local Similarity 98.0%; Pred. No. 1e-138;
Matches 637; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 2525 GGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGTCAAGGACAGAGAAATTTTACCTGG 2584
DB 652 GGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGTCAAGGACAGAGAAATTTTACCTGG 593
QY 2585 AGACACAGGCTGGGAAAGTTGGAGGCCCAAGAACCGAAATCTGAGAGACGAGCTGAGAGAA 2644
DB 592 AGACACAGGCTGGGAAAGTTGGAGGCCCAAGAACCGAAATCTGAGAGACGAGCTGAGAGAA 533
QY 2645 TCAGCCCAAGAGACCAAGTGAACAAGATGGGCTGGGAACTGGAGACCAAGATTTGGGG 2704
DB 532 TCAGCCCAAGAGACCAAGTGAACAAGATGGGCTGGGAACTGGAGACCAAGATTTGGGG 473
QY 2705 AGGTCACTTGAAGCAAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAGCTAC 2764
DB 472 AGGTCACTTGAAGCAAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAGCTAC 413
QY 2765 AGGTCTCCCTGCAAGAGAGAGAGTCAAGTTCAGAGCCCTGCAAGGCTGCAACGGGCGGCC 2824
DB 412 AGGTCTCCCTGCAAGAGAGAGAGTCAAGTTCAGAGCCCTGCAAGGCTGCAACGGGCGGCC 353
QY 2825 TGGAGAGCCAGCTTCGCGCAAGAGAGAGAGAGCTGGAAGAGACCAAGAGAGAGCTGAAG 2884
DB 352 TGGAGAGCCAGCTTCGCGCAAGAGAGAGAGAGCTGGAAGAGACCAAGAGAGAGCTGAAG 293
QY 2885 AGAGATTCAGGACCTTCAAGGACCAATAGAGATGAATTCAGCGCAAAATTTGATGCTTTC 2944
DB 292 AGAGATTCAGGACCTTCAAGGACCAATAGAGATGAATTCAGCGCAAAATTTGATGCTTTC 233
QY 2945 GTAAAGCTGTACTGTAAATCAAGAGAGAGAGAGAGCTTAAACAGCTGACGAGAGACA 3004
DB 232 GTAAAGCTGTACTGTAAATCAAGAGAGAGAGAGCTTAAACAGCTGACGAGAGACA 173
QY 3005 ACGCTGAATCAACAACAACAACTTCTACTTGTCCAAACAACAGTGAAGAGCTTTGGGG 3064
DB 172 ACGCTGAATCAACAACAACAACTTCTACTTGTCCAAACAACAGTGAAGAGCTTTGGGG 113
QY 3065 CCAACAGAGAGATTGTAACAATCGGAGAGAGAGAGTGAAGTGCATCTTCGCGCGGAGAGTCA 3124
DB 112 CCAACAGAGAGATTGTAACAATCGGAGAGAGAGAGTGAAGTGCATCTTCGCGCGGAGAGTCA 53

RESULT 10
BO807302 640 bp mRNA linear EST 31-JUL-2002
LOCUS NISC_KK01b10.y1 NCI CGAP_Brn72 Macaca mulatta cDNA clone
DEFINITION IMAGE5330250.5, mRNA sequence.
ACCESSION BO807302 GI:22031511
VERSION BO807302.1
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecoinae; Macaca.
REFERENCE 1 (bases 1 to 640)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/ILNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
info@image.llnl.gov
Plate: LLAM1838 row: C column: 19
Seq primer: M13RPL reverse primer (ABI).
Location/Qualifiers
1. .640
/organism="Macaca mulatta"
/mol_type="mRNA"
/db_xref="taxon:9544"
/clone_lib="IMAGE:5330250"
/tissue_type="hypothalamus"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI CGAP Brn72"
/note="Organ: brain; Vector: PCMV-SPORE.ccd; Site 1:
NotI; Site 2: EcoRV; Cloned unidirectionally. Primer:
Oligo dt. Average insert size 2.2 kb. Constructed by
Invitrogen. Note: this is a NCI-CGAP Library."

BASE COUNT 166 a 182 c 166 g 126 t
ORIGIN

Query Match 10.2%; Score 628.8; DB 13; Length 640;
Best Local Similarity 98.9%; Pred. No. 1.3e-138;
Matches 633; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 4073 GCGTGTGGCCCGCCATCCAGCCGAGAAAGAGTTCCTCACTCCAGAGAAATTTAGTC 4132
DB 1 GCGTGTGGCCCGCCATCCAGCCGAGAAAGAGTTCCTCACTCCAGAGAAATTTAGTC 60
QY 4133 GGGGTCTTAAAGAACCATGACCAATATTTCTCAACGATTCACAGTGAAGTGAACA 4192
DB 61 GGGGTCTTAAAGAACCATGACCAATATTTCTCAACGATTCACAGTGAAGTGAACA 120
QY 4193 TGGAGCCCAAGAGT 4252
DB 121 TGGAGCCCAAGAGT 180
QY 4253 AATGTCTGAATGACAGT 4312
DB 181 AATGTCTGAATGACAGT 240
QY 4313 GGGGCTTGGCTGTGAATGTGCACACTTTCACGAGGCTTTCGCGGAGAGATGA 4372

Db 241 GCGGCTGCTGCTGCGAATACGACACACTTCACTGAGGCTTCTGCGCGACAAATGA 300
QY 4373 ACTCCAGAGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACCTGMAAGGTGATGA 4432
Db 301 ACTCCAGAGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACCTGMAAGGTGATGA 360
QY 4433 AGGTGCGCAGAAATACAAACGAGAGCAGCAGAGCTGAGCAGAAATGATGCTGCTG 4492
Db 361 AGGTGCGCAGAAATACAAACGAGAGCAGCAGAGCTGAGCAGAAATGATGCTGCTG 420
QY 4493 AGGAGTCAAAAGTCTCATTTATGACAAATGAGCCAGAGCAGCAGAGCTGAGCAG 4552
Db 421 AGGAGTCAAAAGTCTCATTTATGACAAATGAGCCAGAGCAGCAGAGCTGAGCAG 480
QY 4553 AAGAAATTTGAGTCTGCTTCCCGACGAGGATGATATCTATTCATGTCCTGCTGCTT 4612
Db 481 AAGAAATTTGAGTCTGCTTCCCGACGAGGATGATATCTATTCATGTCCTGCTGCTT 540
QY 4613 CCGAATCTCCAAATACAGCCAAAGCAGATGTCCTCATCTACTGAAATGATCTCAC 4672
Db 541 CCGAATCTCCAAATACAGCCAAAGCAGATGTCCTCATCTACTGAAATGATCTCAC 600
QY 4673 CCGACACCACTGCTGCGCCGAGAGACCTCTACTGCT 4712
Db 601 CCGACACCACTGCTGCGCCGAGAGACCTCTACTGCT 640

RESULT 11
Bg976452 881 bp mRNA linear EST 12-JUN-2001
LOCUS 602846269F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4976752 5',
DEFINITION mRNA sequence.
ACCESSION Bg976452
VERSION Bg976452.1 GI:14364089
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 881)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: CGAPsb-rt@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
http://image.llnl.gov
Plate: Llam10970 row: 3 column: 17
High quality sequence stop: 841.
Location/Qualifiers
1..881
source
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976752"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam6"
/note="Torgan: mammary; Vector: pCMV-Sport6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 212 a 269 c 241 g 158 t
ORIGIN

Query Match 9.8%; Score 603.4; DB 12; Length 881;
Best Local Similarity 86.2%; Pred. No. 1.7e-132;
Matches 726; Conservative 0; Mismatches 107; Indels 9; Gaps 5;

QY 5302 AATTACGATCTCTCATTTGAAACCAATTAATTTAGAAATGCG/CATGAAAGCATACG 5361
Db 40 AAGAAAGCATCTCTCATTTGAAACCAATTAATTTAGAAATGCG/CATGAAAGCATACG 99
QY 5362 CTGAGGAATTCCTGATTAAGAAATGACCATTCCTGAGCCTGCTGTGTGCGCTCT 5421
Db 100 CTGATGATGTTCTCTGACAAAGAAAGCAATTCCTGAGCCTGCTGTGTGCGCTCT 159
QY 5422 TCCAAAGCTTCTCTGATTCGAGTGAACAGCGCAGCGCAGCGCAGCGAGATAC 5481
Db 160 TCCAAAGCTTCTCTGATTCGAGTGAACAGCGCAGCGCAGCGCAGCGAGATAC 219
QY 5482 TTGCTGTGCTTCCAGCAATTTGAGAGTTGCTGATTTCTTACGAAAGACGTACCGGACA 5541
Db 220 CTGCTGTGCTTCCAGCAATTTGAGAGTTGCTGATTTCTTACGAAAGACGTACCGGACA 279
QY 5542 GACGATCTCAAGTGAAGTCGCTTACCTTGGCCTTGGCCTTACAGAGAACCTTCTGTTT 5601
Db 280 GATGATCTTAAAGTGAAGTGGCTTACCTTGGCCTTGGCCTTACAGAGAACCTTCTGTTT 339
QY 5602 GTGACCCATTTCACTACTCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5661
Db 340 GTGACCTACTTCACTACTCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 399
QY 5662 CTTGCGCGAGCGATACCTGACATTCGCAACCCCGCTTACCTGAGGCGCTGCTTCTCTCA 5721
Db 400 CTTGCGCGAGCGATACCTGACATTCGCAACCCCGCTTACCTGAGGCGCTGCTTCTCTCC 459
QY 5722 GAGGCGATTTACTTGGCTCTCATACAGAGATTAATTAATTAATTAATTAATTAATTAAT 5781
Db 460 GAGGCGATTTACTTGGCTCTCATACAGAGATTAATTAATTAATTAATTAATTAATTAAT 519
QY 5782 AACCTGTGAAGAGTCCGACCTGAAACACACACCGGCGCCGCTTCCAGTCCGAGCAG 5841
Db 520 AACCTGTGAAGAGTCCGACCTGAAACACACACCGGCGCCGCTTCCAGTCCGAGCAGC 579
QY 5842 CCGAACAAGGAGGCGCCACCAAGTACCAAGAGACATCCAAAGCGCGTGCCTCAGC 5901
Db 580 CCGAACAAGGAGGCGCCACCAAGTACCAAGAGACATCCAAAGCGCGTGCCTCAGC 638
QY 5902 CCGAGCGCGCGCGAAGGCGCCACCAAGTACCAAGAGACATCCAAAGCGCGTGCCTCAGC 5956
Db 639 CCGAGCGCGCGCGAAGGCGCCACCAAGTACCAAGAGACATCCAAAGCGCGTGCCTCAGC 697
QY 5957 ACCGCGAGGAGCGGACCGAGCTCGAGGAGCAAGTCTCTGAGCGCGCGCGCTGAGCGAG 6016
Db 698 ACAGAGAGGAGTCCGAGCAGAGCTCGAGGAGCAAGTCTCTGAGCGCGCGCGCTGAGCGAG 757
QY 6017 AAGAGTCTCCCGCGCGGATCTCAGCAGC-CCGAGAGAGAGGAGTCTCCCGCGCG-AGGCTGTT 6074
Db 758 AAGAGTCTCCCGCGCGGATCTCAGCAGCCTTGAAGAGAGAGGAGTCTCCCGAGGAGAGCTGTT 817
QY 6075 TGAAGACAGCAGCAGGAGCGGCTGCTGCGGAGCGCTGAGAGCCCGGCTGTCCTCAGGT 6134
Db 818 TGAAGACAGCAGCAGGAGCGGCTGCTGCGGAGCGCTGAGAGCCCGGCTGTCCTCAGGT 877
QY 6135 GA 6136
Db 878 AA 879

RESULT 12
B1558919 891 bp mRNA linear EST 05-SEP-2001
LOCUS B1558919
DEFINITION 603240969F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5293913 5',
DEFINITION mRNA sequence.
ACCESSION B1558919
VERSION B1558919.1 GI:15446233
KEYWORDS EST.

Dj	276	TGGAGAGCTACTGATGATGACGAGGAAGAAGCGGAGCAGTGTGCTTGTGAGCTGAGAGA	217
Qy	5063	AAAGTAAACAGCTCCCTGCGCCAGTCCCACTGCTGCGCCAGCCCGAATCTCACCCACAGA	5122
Db	216	AAGTAAACAGTCCCTGCGCCAGTCCCACTGCTGCGCCAGCCCGAATCTCACCCACAGA	157
Qy	5123	TTTTGAAGCTGTCAAGAGGCTGTCACCTGTTTGGGAGCAGCAAGATTGAAACGGAGTC	5181
Dj	156	TTTTGAAGCTGTCAAGAGGCTGTCACCTGTTTGGGAGCAGCAAGATTGAAACGGAGTC	97
Qy	5182	TGCATCTGTGACGCATGCCAGCAAGATGCTATTCTCCGCTACAAAGAAACCTCAGC	5241
Dj	96	TGCATCTGTGACGCATGCCAGCAAGATGCTATTCTCCGCTACAAAGAAACCTCAGC	37
Qy	5242	AAATACCTGCATCCGGAAGAAGATAGAGACCTC	5273
Dj	36	AAATACCTGCATCCGGAAGAAGATAGAGACCTC	5
RESULT 15			
LOCUS	BUI38506	968 bp	mRNA linear EST 25-NOV-2002
DEFINITION	603132657P1 CSEQCHL24 Gallus gallus cdna clone CHEST11b15 5', mRNA sequence.		
ACCESSION	BUI38506		
VERSION	BUI38506.1	GI:25352689	
KEYWORDS	EST.		
SOURCE	Gallus gallus (chicken)		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
AUTHORS	1 (bases 1 to 968) Boagman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Ford, W.T., Tickle, C., Brown, W.R., Wilson, S.A. and Hubbard, S.J. A Comprehensive Collection of Chicken cDNAs Curr. Biol. 12 (22), 1965-1969 (2002)		
TITLE	22335534		
JOURNAL	12445392		
MEDLINE	Contact: Simon Hubbard		
PUBMED	Department of Biomedical Sciences		
COMMENT	University of Manchester Institute of Science and Technology (UMIST)		
FEATURES	1.968		
source	Location/Qualifiers		
	1.968		
	/organism="Gallus gallus"		
	/mol_type="mRNA"		
	/strain="White Leghorn, Hisex"		
	/db_xref="taxon:9031"		
	/clone="CHEST11b15"		
	/dev_stage="16 day embryo"		
	/lab_host="DH10B"		
	/clone_lib="CSEQCHL24"		
	/note="Organ: brain; Vector: plasmid; II KS(+); Site 1:11		
	ECOR; Site 2: NotI; Modification of plasmid; II KS(+)		
	[Stratagene] vector to accommodate cDNA produced with the		
	1-stranded protocol (Construction of uni-directionally		
	cloned cDNA libraries from messenger RNA for improved 3'		
	end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624)		
	624). Cut plasmid; II KS(+); with NotI and EcoRI.		
	Ligate in double stranded adaptor containing BspI and		
	BamHI sites (5'gagcagcgagcccgagatccgagaaag)		
	[5'aattcttttttcgagatccgagcgc]		
BASE COUNT	270 a	206 c	255 g
ORIGIN			237 t
Query Match	9.4%;	Score 581.6;	DB 13; Length 968;
Best Local Similarity	79.5%;	Pred. No.2.9e-127;	
Matches /63; Conservative	0;	Mismatches 189;	Indels 8; Gaps 6

Sequence 12, Application US/08422699A
Patent No. 5955265
GENERAL INFORMATION:
APPLICANT: Brook, J. David
APPLICANT: Housman, David E.
APPLICANT: Shaw, Duncan J.
APPLICANT: Harley, Helen G.
APPLICANT: Johnson, Keith J.
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02713
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/422,699A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/422,706
FILING DATE:
APPLICATION NUMBER: US 08/023,612
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/839,255
FILING DATE: 20-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/01545
FILING DATE: 19-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00253
FILING DATE: 05-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9202485.0
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-5830A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2726 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-422-699A-12

Query Match 3.7%; Score 227.2; DB 2; Length 2726;
Best Local Similarity 54.1%; Pred. No. 5,1e-49;
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 3;

QY 194 CTGCTCTGATGAAGATTAAAGCAGTGAAGCACTTTGTCGGAAGTATTCGACACCATAG 253
DB 243 CGGAAGTGGCCAGAGCAAGTACGTGGCCGACTTTGTCAGTGGGCGAGCCATCGTGG 302
QY 254 CTGAATTACAGGAGCTCCAGCCTTGGGCAAGAGCTTGAAGTCAAGAACTCTTGAAGTT 313
DB 303 TGAGGCTTAAGAGGCTCGACTGCAAGAGGAGCACTTCGAGATTTGAAAGGATCGAGC 362
QY 314 GTGTCACTTGTGCTGAAGTGAAGTGTGTAAGAGAAAGCAACGCGGACATCTATGTCTA 373

DB 363 GCGGGGCGTTTCAGCGAGGTAGCGGTATGAAGATGAGCAGACGGGCGAGGTATGCCA 422
QY 374 TGAAGATGATGAAGAGAGAGCTTTATTTGGCCAGAGAGAGGTTTCAATTTTGGAGAG 433
DB 423 TGAAGATCATGAAGAGAGAGAGTGTCTAAAGAGGGGCGAGGTGTCTGCTTCGTAAG 482
QY 434 AGCGGAATATTTATCTCGAAGACACAGCCCGTGTATCCCAATTAACATATGCTTTC 493
DB 483 AGAGGACCTGTGTGTGAATGAGGACCCGGCGGTGATCAACGACGTGCACTTCGCTTCC 542
QY 494 AGAAGAAAATCAACCTTTATCTGTGATGAAGTATCAAGCTGGAGGGGAGCTTGTCTAC 553
DB 543 AGAATGAGACTTACCTGTACCTGTGTATGAGATTAATCTGGCGGGAGACTGTACAC 602
QY 554 TTTGAAATGATATGATGAGACCACTTATGATGAAAACCTGATTAAGTTTACCTAGTACG 613
DB 603 TGCTGAGCAAGTTTGGGGAGCGGATTCGCGCGAGATGCGCGCTTCTACCTGGCGAG 662
QY 614 TGAATTTGGCTGTTACAGAGCTTCACTGTATGAGATAGTGCATCGACATCAAGCTTG 673
DB 663 TTGTATGCGCATGACCTGCTGACACCGGCTTGGCTTACCTGACAGAGCATCAACCG 722
QY 674 AGAATCTCTGTTGACCGCAGAGACATCAAGCTGTGATTTGATTCGCGCGA 723
DB 723 ACAATCTCTGCTGAGACCGCTGTGGCCATCTCGCTGCGCACTTCGCTCTCA 782
QY 734 AAATGAATTCAAACAGATGTGAATGCCAACTCCCGATTTGGAGCCCAAGATTACATGG 793
DB 783 AGCTGGGGCAGATGAGAACGTTGCGGTGCTGTGTGGCTGTGGGCACTCCAGACTACCTG 842
QY 794 CTCTGAAGTCTGA--CTGTGATGAACGCGGATGGAAGAGGACCTTACGCGCTGAGCT 850
DB 843 CCCCCGAGATCTGAGAGCTGTGGGCGTGGGCTTGGGACAGGCACTACGCGCCGAGT 902
QY 851 GTGACTGATGTCAGTGGGCGTGTGATTTGCTATGAGATGATTTATGGGAGATCCCGCTTG 910
DB 903 GTGACTGTGGGCGCTGTGTGTATTCGCTATGAATGTTCTATGGGACAGACCCCTTCT 962
QY 911 CAGAGGGAACCTCTGCGCAGAACCTTCATTAATTAATTAATTTCCAGCGGTTTGAAT 970
DB 963 AGCGGATTCAGACGGGGAGACCTTATGCAAGATGCTCCATCAAGAGACACTCTCTC 1022
QY 971 TTCCAGATGACCCCAAGTGAAGT--GACTTTTGTATCTGATTCAAAGCTTGTGT 1027
DB 1023 TCCCGCTGTGACCAAGGGGTCCCTGAGAGGCTCGAACAATTCATTCAGCGGCTTCTGT 1082
QY 1028 GCGGCGAAGAAAGAGACTGAAGTTGAAGTCTTTGCTGC-----CATCTTTCT 1078
DB 1083 GTCCCCCGAGAGACAGGCTGGGCGGGGTGGAGCAGGCACTTCGAGACATCTCTCT 1142
QY 1079 TCTTAAATTTGACTGGAACAACTTCGTACTCTTCTCCCTTGTGTTCCACCCCTCA 1138
DB 1143 TCTTTGCTCTGATGGATGTGTCTCCGGGACAGGCTGCCCTTTTACACCGGATTTGG 1202
QY 1139 AGCTGAGATGACACCTTCATTTTGA 1166
DB 1203 AAGGTGCCACCGACATGCAACTTGA 1230

RESULT 4
US-08-422-706B-12
Sequence 12, Application US/08422706B
Patent No. 5977333
GENERAL INFORMATION:
APPLICANT: Brook, J. David
APPLICANT: Housman, David E.
APPLICANT: Shaw, Duncan J.
APPLICANT: Harley, Helen G.
APPLICANT: Johnson, Keith J.
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF
NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Millitia Drive
;; CITY: Lexington
;; STATE: Massachusetts
;; COUNTRY: US
;; ZIP: 02713
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/422,706B
;; FILING DATE: 14-APR-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/284,543
;; FILING DATE: 08-AUG-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/023,612
;; FILING DATE: 26-FEB-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/839,255
;; FILING DATE: 20-FEB-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/01545
;; FILING DATE: 19-FEB-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/GB93/00253
;; FILING DATE: 05-FEB-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: GB9202485.0
;; FILING DATE: 06-FEB-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: MIT-5830A2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-861-6240
;; TELEFAX: 617-861-9540
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2726 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-08-422-706B-12

Query Match 3.7%; Score 227.2; DB 2; Length 2726;
Best Local Similarity 54.1%; Pred. No. 5,1e-49;
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 3;

QY 194 CTGCTCTGATGAAGATTAAAGCAGCGTAGCACTTTGTCGGGAAGATTCCGACACCATAG 253
DB 243 CCGAAGTGGCCAGAGCAAGTACGTGGCCGACTTCTTGACATGGGGAGCCCATCGTGG 302
QY 254 CTGAGTTACAGAGGCTCCAGCCTTCGGCAAGAGCTTCGAACTCAGAACTCTTGTAGTT 313
DB 303 TGAAGCTTAAGAGAGTCCGACATGCAGAGGAGCACTTGCAATTCGAAAGTGATCGGAC 362
QY 314 GTGTGTCATCTTCTGTAAGTGCAGGTGGTAAAGAGAAAGCAACCGGGACATCTATGCTA 373
DB 363 GGGGGGCGGTTCAGCAGAGGTAGCGGTAGTGAAGATAGACAGCGGCGCAGGTATGCGCA 422
QY 374 TGAAGTGTATGAAGAAGAGCTTATTGGCCAGAGCAGGTTTCATTTTGAAGAAG 433
DB 423 TGAAGATCATGAACAAGTGGAGCATGCTAAGAAGGGGCGAGGTGCTTCCGTGAGG 482
QY 434 AGCGGAACATATATCTCGAAGACACAAGCCGTTGATCCCAATTAAGTATGCTTTC 493
DB 483 AGAGGAGCGTGTGTGAATGGGAGCCGGCGGTGATCAGCAGCTGCACCTTCCTTCC 542

QY 494 AGGACAAAATACCTTTATCTGTATGTGAATATCAGCTGAGGGGACTTGTCTGAC 553
DB 543 AGATGAGAACTACTGTAAGTCTGTATGAGGTATTAAGTGGGGGAGCCTGCTGACAC 602
QY 554 TTTTGATATGATAGAGGACCAAGTTAGTAAAACTGATACAGTTTACTACTAGGC 613
DB 603 TGCTGACCAAGTTTGGGAGAGGAGTCCGGCGAGATGGCGGCTTCTACTGGGGGGA 662
QY 614 TGATTTTGGCTGTTCACAGCGTTTCATCTGATGGGATACGTGATCGAGACATCAAGCTG 673
DB 663 TTGTCAATGCGCATAGATCTGGTGTGACACCGGCTTGGCTAGCTGACAGGACATCAACCG 722
QY 674 AGAATTTCTGTTGACCGCACAGAGACATCAAGCTGTGGATTTTGGATTTGCGCGGA 733
DB 723 ACAATCTCTGCTGACCGCTGTGGCCACATCCGCTGGCCGACTTGGCTTCTGCTCA 782
QY 734 AATGATTTCAACAAGATGTGAATGCGAACTCCGATTTGGGACCCCAATTACATG 793
DB 783 AGTCCGGGCGAGATGGAACGGTGGCGGTGCTGTGGCTGTGGGACCCCAACTACCTGT 842
QY 794 CTCTGAAGTCTGA---CTGTGATGAAAGGGGATGAAAGGCACTACGCTGAGCT 850
DB 843 CCCCAGATCTCTCAGGCTGTGGGGCGGTGGGCTTGGACAGGACGCTACGGGCCGAGT 902
QY 851 GTGACTGTGTCAATGAGGCGCTGATGCTTATGAGATGATTTATGGAGATCCCTTTCG 910
DB 903 GTGACTGTGGGCGCTGGGTGATTCGCTTATGAATGTTCTATGGGACAGCGCTTCT 962
QY 911 CAGAGGGAACTCTGCCAGCACTTCATTAATTAATTTCCAGGGTTTTGGAAT 970
DB 963 ACGGGATTTCCAGCGCGAGACCTTAGGCAAGATGTCACATCAAGAGACCTCTCTC 1022
QY 971 TTCCAGATGACCCCAAGTGAAGCT---GACTTCTTATCTGATTCGATTAAGCTTGTGT 1027
DB 1023 TGCCGCTGTGAGCAAGAGGCTCCTTAGAGAGGCTCGAGACTTATAGCGGTGTGCT 1082
QY 1028 GCGGCGAAGAAAGAGACTGAAGTTGAAGTCTTTGCTG-----CATCTTTCT 1078
DB 1083 GTCCCGCGAGACACGCTGGGCGGGGTGAGACAGCGACTTCGAGCACATCCCTTCT 1142
QY 1079 TCTCTAAATTTGACTGGAACAACATTTGTAATCTCTCCCCCTCGTTCACACCTCA 1138
DB 1143 TCTTTGGCTGACTGGGATGCTCTCCGGGACAGCTGCCCCCTTTTACACCGGATTTG 1202
QY 1139 AGTCTGAGATGACACTCCCAATTTTGA 1166
DB 1203 AAGTGCAACCGACATGCAACTTCA 1230

RESULT 5
US-08-422-699A-8
; Sequence 8, Application US/08422699A
; Patent No. 5955265
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/839,255
FILING DATE: 20-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/01545
FILING DATE: 19-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00253
FILING DATE: 05-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB9202485.0
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-5830A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1746
US-08-422-706B-8

Query Match 3.6%; Score 220.4; DB 2; Length 2511;
Best Local Similarity 54.6%; Pred. No. 2.9e-47;
Matches 512; Conservative 0; Mismatches 411; Indels 15; Gaps 3;

QY 244 GACACCAATGCTGAGCTTCCAGAGCTTGGCAAGAGCTTGGAGTCAGAGT 303
DB 43 GCCATCGGTGAGGCTTAAAGAGCTCCAGTCGACAGAGGAGCACTTGAAGTCTGAAG 102
QY 304 CTGTAGGTGAGTCACTTGTCTGAAGTGCAGGTGTAAGAGAGCAACCGGGGAC 363
DB 103 GTGATCGACCGGGGCGTTCCAGCAGAGTAGCGGTAGTAAGATGAACAGCGGGCAG 162
QY 364 ATCTATGCTATGAAGATGATGAAGAGCTTATTGGCCAGAGCAGAGTTCATT 423
DB 163 GTGATGCCATGAAGATCATGAACAGTGGGCACTGAAGAGGGGAGAGGTGTCGAC 222
QY 424 TTGAGAGAGCGGCAATATATCTGGAAGCACAAGCCGTGATCCCCAATTACG 483
DB 223 TTCCGTGAGAGAGGAGCGTGTGTGAATGGGACCGCGGTGATCACAGCTGCAC 282
QY 484 TATGCTTTGAGCAAAAATCACCTTATCTGTCATGATATCAGCTGAGAGGAG 543
DB 283 TTCCCTTCCAGATGAGAACTACTCTGATCTGTCATGATATTAAGTGGCGGGAC 342
QY 544 TTGCTGTCACTTTGAATGATATGAGGACCAATTAGTAAACCTGATACAGTTTAC 603
DB 343 CTGCTGACACTGTGAGCAAGTTGGGAGCGGATTCGCGCGAGATGGCGCTTAC 402
QY 604 CTAGCTGAGCTGATTTTGGTGTTCACAGCGTTCATCTGATGGGATTAAGTCACTGAGAC 663
DB 403 CTGGGAGGATTTGTATGCGCATGACTCGGTGACCGGCTTGGTGTACGTCAAGGGAC 462
QY 664 ATCAAGCTGAGAACATTTCTGTGACCGCAGAGACACATCAAGCTGTGATTTTGA 723
DB 463 ATCAAGCCGACACATCTGTGTGACCGCTGTGGCCACATTCGCTTGGCGACTTGGC 522
QY 724 TCTGCGGGAATGATTAACAAGATGTAATGCAAACTCCGATTTGGACCCCA 783
DB 523 TCTTGCTCAAGCTGCGGAGATGGAACGGTGCCTGTGTGGCTGTGGGACCCCA 582
QY 784 GATTACATGGCTCTGGA--AGTGTGCTGTGATGAACGGGATGGAAGGACCTAC 840

DB 583 GACTACTGTCCCGGAGATCTGCAGCTGTGGCGGTGGGCTGGACAGCAGCTAC 642
QY 841 GGCGTGAAGTGAAGTGTGTGATGAGCGGTGATGCTATGATATTAATGAGAGA 900
DB 643 GGGCCGAGTGTGATGTGGCGGTGTGTATTCGCTATGAATGTTCTATGGCAG 702
QY 901 TCCCGCTTGGAGAGGAACTTGGCCAGAACTTCAATTAATTAATTAATTTCCAGCG 960
DB 703 AGCGCTTCTACGGGATTCACCGCGGAGACCTATGGCAAGTGTCTACAGAGAG 762
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGCT--GACTTCTGATCTGATTCAA 1017
DB 763 CACCTCTCTCGCGGTGTGAGAGAGGGTCCCTGAGAGGCTCGAGACTTCAATTCAG 822
QY 1018 AGCTTTGTGCGGCGCGAAGAGAGATGTAAGTTGAAGTCTTGTCTG----- 1068
DB 823 CGGTGTGTGTCCCGGAGACAGGCTGTGGCGGGGTGAGAGCGAGACTTCCGAGCA 882
QY 1069 CATCCTTCTCTCTAAATTAATGACTGGAACAATTCGTAATCTCTCCCGCTTCTGTT 1128
DB 883 CATCCCTTCTCTTGTGGCTGACTGAGTGTGTCTCCGGACAGGCTGCCCTTTACA 942
QY 1129 CCACCTCAAGCTGACATGACACTTCAATTTGA 1166
DB 943 CCGATTTCGAAGGTGCCACGACACATGCAACTTCGA 980

RESULT 7
US-08-484-044-11
Sequence 11, Application US/08484044
GENERAL INFORMATION:
APPLICANT: Caskey, C. T.
APPLICANT: Fu, Ying-Hui.
APPLICANT: Friedman, David L.
APPLICANT: Pizutti, Antonio
APPLICANT: Fenwick, Raymond G.
TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSER: Fulbright & Jaworski, L.L.P.
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,044
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/019,940
FILING DATE: 19-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Paul, Thomas D.
REGISTRATION NUMBER: 32,714
REFERENCE/DOCKET NUMBER: D-5443
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713/651-5325
TELEFAX: 713/651-5246
TELEX: 762829
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 3182 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-08-484-044-11

Query Match 3.6%; Score 220; DB 1; Length 3182;
Best Local Similarity 54.6%; Pred. No. 4.2e-47;
Matches 511; Conservative 0; Mismatches 410; Indels 15; Gaps 3;

QY 246 CACCATAGCTGAGTTCACAGAGCTCCAGCCTTCGGCAAGAGCTTCGAATCGAAGCTC 305
DB 743 CATCGTGTGAGGCTTAAAGAGGCTCCAGCTGACAGAGGAGGAGCTTCAGATTCCTGAAGGT 802
QY 306 TGTAGGTGTGTCACTTGTCTGAAGTGCAGGTGTGTAAGAGAGAAAGCAACGGGAGCAT 365
DB 803 GATCGAGCGGGGGCGGTTCAGCGAGGTAGGGGTAGTGAATGAGAGACGGGCCAGGT 862
QY 366 CTATGCTATGAAGATGATGAAGAGAGAGGCTTTATTTGGCCAGAGAGGTTTCATTTT 425
DB 863 GTATGCCATGAAATGATGAAGAGAGAGGTCATGTAAGAGAGGAGGAGGTGCTGTCCTT 922
QY 426 TGAAGAGAGCGGAGCAATATATCTGAAAGCAAGAGCCCTGATCCCAATTAAGTA 485
DB 923 CCGTGAAGAGAGAGAGGTGTGTGTAATGGAGACCGGCGGTGATACGAGCTGCACCTT 982
QY 486 TGCCTTTCAGAGCAAAATACCTTTATCTGTATGATGAAATATCAGCCTGAGGGGACTT 545
DB 983 CGCCTTCCAGGATGAGAGACTACTGTATCCTGTCAATGAGATATTAAGTGGCGGAGCCT 1042
QY 546 GCTGTCACTTTTGAATGATATGAGAGCAAGTTAGATGAAAACCTGATACAGTTTAACT 605
DB 1043 GCTGACACTGTGAGAGAGTTTGGGAGCGGATTTCCGCGCGAGATGCGGCTTCACTT 1102
QY 606 AGCTGAGCTGATTTTGGCTGTTCACAGCGCTTCATCTGATGAGATGAGTATGAGAGCAT 665
DB 1103 GGGGAGATTTGATGAGGCTTATGAGCTGAGCTGAGAGCGGCTTGGCTGAGAGAGCAT 1162
QY 666 CAAGCTGAGAGCAATTTCTGTTTGAACCGGACAGAGACATCAAGCTGTGATTTTGGATC 725
DB 1163 CAAGCCGAGCAATCTGCTGTGAGCGGCTGTGAGCAATCCGCTGAGCGAGCTTCCGCTC 1222
QY 726 TGGCGGAGAAATGATTTCAACAGAGATGTGAATGCAAACTCCGATTTGGAGCCCGAGA 785
DB 1223 TTGCCTCAAGCTGCGGAGAGATGAAACGGGTGCTGTGGGCTGTGGGACCCGAGC 1282
QY 786 TTACATGAGCTCCTGCA---AGTGTGATGATGAAACGGGAGTGAAGAGGACCTAACGG 842
DB 1283 CTACCTGTCCCCGAGATCTGCAAGCTGTGGCGGCTGGGCTTGGAGCAAGGACGCTAACGG 1342
QY 843 CCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 902
DB 1343 GCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1402
QY 903 CCGCTTGCAGAGAGGAACTCTGCGCAAGACCTTCAATTAATGATTAATTTCCAGCGGTT 962
DB 1403 GCGCTTCTAAGCGGATTCACAGCGGAGACCTATGCAAGATGCTCACATCAAGAGAGA 1462
QY 963 TTGAAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1019
DB 1463 CCTCTCTGCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1522
QY 1020 CTGTGTTGTCGCGCAGAGAGAGAGAGTGAAGTTTGAAGTCTTTTGTGTC-----CA 1070
DB 1523 GTTGTGTGTGTCGCGCAGAGAGAGAGAGTGAAGTCTTTTGTGTC-----CA 1582
QY 1071 TCGTTTCTTCTCTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1130
DB 1583 TCGCTTCTTCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1642
QY 1131 CACCCCTCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1166
DB 1643 GGATTTGGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1678

RESULT 8
US-09-804-471A-3

; Sequence 3, Application US/09804471A
; Patent No. 6479269
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001164
; CURRENT APPLICATION NUMBER: US/09/804,471A
; CURRENT FILING DATE: 2001-03-13
; NUMBER OF SEQ. ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 3
; LENGTH: 174493
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(174493)
; OTHER INFORMATION: n = A,T,C or G
US-09-804-471A-3

Query Match 3.3%; Score 205; DB 4; Length 174493;
Best Local Similarity 100.0%; Pred. No. 3.2e-42;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 753 GGTGAATGCCAACTCCGATTTGGAGCCCGAGATTACATGCTCTGTAAGTGTGACTGT 812
DB 130289 GGTGAATGCCAACTCCGATTTGGAGCCCGAGATTACATGCTCTGTAAGTGTGACTGT 130348
QY 813 GATGAACGGGAGATGAGAAAGGACCTTACGCGCTGAGCTGTGATGCTGTGATGAGTGGCCT 872
DB 130349 GATGAACGGGAGATGAGAAAGGACCTTACGCGCTGAGCTGTGATGCTGTGATGAGTGGCCT 130408
QY 873 GATTGCTATGAGATGATTTATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 932
DB 130409 GATTGCTATGAGATGATTTATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 130468
QY 933 CTTCATTAACATTATGATTTTCCAG 957
DB 130469 CTTCATTAACATTATGATTTTCCAG 130493

RESULT 9
US-08-630-822A-61
; Sequence 61, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLLENFELDS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020

```

; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5..2706
; US-08-630-822A-61

```

Query Match 3.2%; Score 195.6; DB 2; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41; Mismatches 384; Indels 27; Gaps 4;

```

Matches 493; Conservative 0;
QY 281 CAAGAGCTTCGAGTCAGAGTCTTGAGTGTGTCACCTTGTCTGAAGTCAGAGTGG 340
DB 78 CAGATGATTTTAAATTAATAAGTATTGTCGAGAGACATTTGGTGAAGTACAGTTAG 137
QY 341 TAAGAGAGAAAGCAACCGGGGACATCTATGTAAGTGAAGGAAGAGCCTTAT 400
DB 138 TCGACACAAATCAATCGACAAAGTTTGTATGAACGCCCTATCAAAATTTGAATGA 197
QY 401 TGGCCAGAGAGCAGGTTTCATTTTGAAGAGCGGAACATATTATCTCGAAGACAA 460
DB 198 TTAAGAGACACAGACTGTCATTTTGGGAGAGACGTATATATGCTCATGCAAAAT 257
QY 461 GCCCGTGAATCCCCAATTACAGTATCCCTTCAGAGCAAAATCACTTATCTGTGCA 520
DB 258 CAGATGATGTGACATTTACATTTTGGCTTTCAAGATCAAAATCTTTATATGTCGA 317
QY 521 TGGAAATGAGCTGAGAGGGGACTTGTCTGACCTTTGTAATGATGAGACCAATTAG 580
DB 318 TGGATTATATGCCGGGGGAGTGTGAGTGTATG-----TCCGATTATGAATTC 371
QY 581 ATGAAACCTGATACAGTTTACCTTACCTGAGCTGATTTTGGCTGTTCACAGGCTCATC 640
DB 372 CAGAAAAATGGGCAATGTTCTATACATAGAGAGTGTGCTAGACATTTGATACATTCAC 431
QY 641 TGATGGATACGTGATGACATGACATCAAGCTTGAGAAATTCCTGTTGACCGCAGAGAC 700
DB 432 CATGGGATTTGATACATGATGATTAACCTGATATATGCTTCTAGACAAATATGATC 491
QY 701 ACATCAAGCTGTGATTTTGGATCTGCCCGGAAATGAAATTAACAAGATGAGTGAAG 760
DB 492 ATTTAAAGTTAGCTGCTTGGAACTGTATGAAATGATACAGATGTTGGTACGTT 551
QY 761 CCAAACTCCCGATTGGGAGCCCGCAGATTACATGAGCTCTGAGTGTGATGATGAAG 820
DB 552 CTATATATGCTGTGGAACGCTGATTAATTTCTCCGAGT-----TTTGCAATCCC 605
QY 821 GGGATGAAAAAGGACCTTACGCTGAGCTGTGATGCTGTGATGAGGCGGTATGCTCCT 880
DB 606 AAGGTGTGAAAGAGTTTACGCTGCGAATGCGATTGGTGTCTGTGGGAAATTTTGTG 665
QY 881 ATGAGATGATTTATGGGAGATCCCTCTGCGAGAGGGAACCTCTGCACAGACCTTCAAT 940
DB 666 ATGAAATGTTATTTGAGAAACACCTTTTATGCAAGACAGTGTGTTGGAACTTACAGTA 725
QY 941 ACATTAATGAATTTCCAGCGGTTTGAATTTCCAGATGAACCCCAAGTGAAGCAGTACT 1000
DB 726 AATATATGATTCAGAAACTCATTAATTTCTCCCAAGAGTGAATTAAGCAATATAG 785
QY 1001 TTCTTAATCTGATTCAGAGCTTGTGTGGGCGCAGAAAG-----AGACTGAAGT 1051
DB 786 CCGGATCTTTGATACAGAGATTTTAAACAGACAGAAACACAGCGCTTTAGGCGAATGAAG 845
QY 1052 TTGAAGGTCTTGCTCCATCCTTTCTCTCTAAATGATGATGA-----ACAACTTC 1105

```

```

DB 846 TGGAGAAATTAAGACATCATTTTCTAATAAATGATCAATGACTTTTGAACAATTTAA 905
QY 1106 GTAACCTCTCCCTCCCTTGTGTTCCACCCCTCAAGTGTGACATGACACCTCAATTTG 1165
DB 906 GAGACTCTGCCCTCCTGATGTCAGAGCTGATGATGATATGATACAGAACTTTG 965
QY 1166 ATGA 1169
DB 966 ATGA 969

```

RESULT 10

US-09-005-069-61
Sequence 61, Application US/09005069
Patent No. 5932470

GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,069
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 2706 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2706
US-09-005-069-61

```

Query Match 3.2%; Score 195.6; DB 2; Length 2706;
Best Local Similarity 54.5%; Pred. No. 9e-41;
Matches 493; Conservative 0; Mismatches 384; Indels 27; Gaps 4;
QY 281 CAAGAGCTTCGAGTCAGAGTCTTGAGTGTGTCACCTTGTCTGAAGTCAGAGTGG 340
DB 78 CAGATGATTTTAAATTAATAAGTATTGTCGAGAGACATTTGGTGAAGTACAGTTAG 137
QY 341 TAAGAGAGAAAGCAACCGGGGACATCTATGTAAGTGAAGGAAGAGCCTTAT 400
DB 138 TCGACACAAATCAATCGACAAAGTTTGTCTATGAAAGCGCTTATCAAAATTTGAATGA 197

```

CITY: DEN

22

Query Match	3.2%	Score 195.6	DB 4	Length 2706
Best Local Similarity	54.5%	Pred. No. 9e-41		
Matches 493	Conservative 0	Mismatches 384	Indels 27	Gaps 4
QY	281	CAAAAGCCTTGAAGTCAGAAGCTTGTAGGTGTGTCACCTTGTGTAAGTCAGAGTGC	340	
Db	78	CAGATGATTTAAATTTAAATAAAGTTATTTGGTCGAGGAGCATTGGTGAAGTACAGTTAG	137	
QY	341	TAAGAAGAAAGCAACCGGGGACATCTATGCTATCAAAATGATGTAAGAAAGAGGCTTTAT	400	
Db	138	TGCACACAAATCAACCTGCACAGTTTGTGCTATGAACGCTATCAAAATTTGAANAATGA	197	
QY	401	TGGCCAGAGACAGGTTTCATTTTTTGAAGAAGCGGACATATTATCTCGAGACACAA	460	
Db	198	TTAAGAGACGACCTCTGCATTTTTTTTGGAAAGAACGTATATATGCTCATGCAAAAT	257	
QY	461	GCCCGTGGATCCCCCAATTAAGATGCTCTTCAGACAAAAATACCTTTATCTGTGTA	520	
Db	258	CAGATGGATGTGACAAATTTACATTTTGTCTTTCAGATCAAAATAATCTTTATATAGTCA	317	
QY	521	TGGAATATCAGCCCTGAGAGGACTTGCTGCATCTTGAATATGATATAGAGCAGTTGA	580	
Db	318	TGGATTTATATCCGGGGGGTGACTTGGTGAAGCTTTATG-----TCCGATTTATGAATTC	371	
QY	581	ATGAAAACCTGTATACAGTTTAACTAGCTGACTGATTTGGCTGTTCAAGCGTTTATC	640	
Db	372	CAGAAAATATGGGCATGTTCTATACAAATGAAAGGTGTCATGCACCTTGATATCAATTCAC	431	
QY	641	TGATGGGATACGTGCATCGAGACATCAAGCTGAGAACATTCCTGTTGACCGACAGGAC	700	
Db	432	CCATGGGATTTGTATCATCTGTGATGTAAACCTGTATATATGCTTTACAGCAAAATATGTC	491	
QY	701	ACATCAAGCTGTGTGATTTTGGATCTGTGCGCGGAAAATGAATTCAAAACAGATGTGTAATG	760	
Db	492	ATTTAAAGTTAAGCTGACCTTTGGAACTGTATGAAAATATGATACAGATGGTTTGGTACGTT	551	
QY	761	CCAAATCCCGATTTGGAGCCCAATTAACATGAGCTCCCTGAAGTGTGACTGTGATGAAG	820	
Db	552	CTAATATATGCTTTGGAACGCGCTATATCAATTTCTCCGAAGT-----TTTGCAGTCCC	605	
QY	821	GGGATGGAAGGACCTAGGCGCTGGACTGTGACTGTGTCACTGGGCGTGATTTGCTT	880	

Db 606 AAGGTGTGAAGAGTTACGTCGTGAATGCGATTGTGTCTGTGGAAATTTTGT 665
QY 881 ATGATGATGATTTATGAGAGATCCCTTCGACAGAGGAACTCTGCCAATCTTCAATA 940
Db 666 ATGAATGATTTATGAGAGAACACCTTTTATGACAGACGTTGTGGAACCTTACAGTA 725
QY 941 ACATTATGATTTCCAGCGGTTTTGAAATTTCCAGATGACCCCAAGTACAGTACT 1000
Db 726 AATTATGATCAGAAACTCATTTACTTCTCCAGAAAGTGAATTAAGCCATATAG 785
QY 1001 TTCTTGATCTGATTCAGAGCTTGTGTGGGCCAGAAAG-----ACACTGAAT 1051
Db 786 CCGGATCTTTGATACAGGATTTTAAACAGACAGAACACAGGCTTTAGGCAAAATGAAG 845
QY 1052 TTGAAGTCTTGTGCGCATCTTCTCTCTAAATGACTGSA-----ACAACATTC 1105
Db 846 TGGAGAAATTAAGACACATTCATTTTATTAATGATCAATGACTTTTGACAATTTAA 905
QY 1106 GTAATCTCTCCCTCCCTTCGTTCCACCCCTCAAGCTGACATGACACCTCCAAATTTG 1165
Db 906 GAGACTCTGCCACCTGATGACGACAGCTGATGATGATGATGATGATGATGATGATGATG 965
QY 1166 ATGA 1169
Db 966 ATGA 969

RESULT 12
US-09-004-730A-20

; Sequence 20, Application US/09004730A
; Patent No. 6485968
; GENERAL INFORMATION:
; APPLICANT: Weber, Eric
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Sim, Gek-Kee
; APPLICANT: Frank, Glenn
; APPLICANT: Wallenfels, Lynda
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUC
; FILE REFERENCE: 2618-17-C5-PUS-1
; CURRENT APPLICATION NUMBER: US/09/004,730A
; PRIOR FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: PCT/97US/18669
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 20
; LENGTH: 2706
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)..(2704)
US-09-004-730A-20

Query Match 3.2%; Score 195.6; DB 4; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41; Matches 493; Conservative 0; Mismatches 384; Indels 27; Gaps 4;

QY 281 CAAAGACTTCAGAGTCAGAGCTTGTAGGTGTGTGCTACTTGTCTGAAGTCAGAGTGG 340
Db 78 CAGATGATTTTAAATTAATAAGTATGTTGTGAGACATTTGTGAGTACAGTTAG 137
QY 341 TAAGAGAGAAAGCAACCGGAGACATCTATGCTATGTAAGTGAAGAAAGGCTTTAT 400
Db 138 TCGACACAAATCAATGACAGATTTTGTCTATGAAAGCCATCAAAATTTGAATGA 197
QY 401 TGGCCAGAGAGAGAGTTTATTTTGGAGAGAGGGAACATATATCTCGAAGACAA 460
Db 198 TTAAGAGACAGACTCTGATTTTGTGAGAGAGAGTATATATATGCTCATGCAAAAT 257
QY 461 GCCCGTGAATCCCAATTAACAGTATGCTTTTCAGAGCAAAATCACTTTATCTGTGA 520
Db 258 CAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 317

QY 521 TGAATATCAGCTTGAGAGGAGCTTGTGCTGACATTTTGAATATGATATGAGACAGTTAG 580
Db 318 TGAATATATGCGGAGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 371
QY 581 ATGAATACCTGATATACATTTTACCTAGCTGAGCTGATTTTGGCTGTTCACAGCTTATC 640
Db 372 CAGAAATATGAGCAATGCTTATACATATGAAAGTGTGTACACCTTGATACAAATTCAC 431
QY 641 TGAATGATAGCTGATGAGACATCAAGCCGAGAACCTTCGTTGCGCCAGACAGAC 700
Db 432 CCATGAGATTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 491
QY 701 ACATCAAGCTGTGATGATTTTGTGATCTGCGCGCAAAATGATTAACAAAGATGTGATG 760
Db 492 ATTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 551
QY 761 CCAAACTCCGATTTGAGAGCCCAATTAATGATGATGATGATGATGATGATGATGATG 820
Db 552 CTAAATATGCTGTTGAGACGCTGATTAATTAATTTCTCCGAAATG-----TTTGCAGTCCC 605
QY 821 GGGATGAAAAGCACTACGCTGAGCTGATGATGATGATGATGATGATGATGATGATG 880
Db 606 AAGGTGTGAAGAGCTTTACGCTGTAATGCGATTTGTGTGTGTAATTTTGT 665
QY 881 ATGATGATGATTTATGAGAGATCCCTTCGACAGAGGAACTCTGCCAATCTTCAATA 940
Db 666 ATGAATGATTTATGAGAGAACACCTTTTATGACAGACGTTGTGTAATTTACAGTA 725
QY 941 ACATTATGATTTCCAGCGGTTTTGAAATTTCCAGATGACCCCAAGTACAGTACT 1000
Db 726 AATTATGATCAGAAACTCATTTACTTCTCCAGAAAGTGAATTAAGCAATATAG 785
QY 1001 TTCTTGATCTGATTCAGAGCTTGTGTGGGCCAGAAAG-----ACACTGAAT 1051
Db 786 CCGGATCTTTGATACAGGATTTTAAACAGACAGAACACGCTTTAGGCAAAATGAAG 845
QY 1052 TTGAAGTCTTGTGCGCATCTTCTCTCTAAATGACTGSA-----ACAACATTC 1105
Db 846 TGGAGAAATTAAGACACATTCATTTTATTAATGATCAATGACTTTTGACAATTTAA 905
QY 1106 GTAATCTCTCCCTCCCTTCGTTCCACCCCTCAAGCTGACATGACACCTCCAAATTTG 1165
Db 906 GAGACTCTGCCACCTGATGACGACAGCTGATGATGATGATGATGATGATGATGATG 965
QY 1166 ATGA 1169
Db 966 ATGA 969

RESULT 13

US-08-981-799A-20
; Sequence 20, Application US/08981799A
; Patent No. 6576238

; GENERAL INFORMATION:
; APPLICANT: Weber, Eric
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Sim, Gek-Kee
; APPLICANT: Frank, Glenn
; APPLICANT: Wallenfels, Lynda
; TITLE OF INVENTION: "NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT S
; FILE REFERENCE: 2618-17-C5-PUS
; CURRENT APPLICATION NUMBER: US/08/981,799A
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: PCT/97/18669
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 20
; LENGTH: 2706
; TYPE: DNA
; ORGANISM: Ctenocephalides felis

198 GAACATAGATATATTTCTTAATATAGATAGAAAAATTTGAAAAATCAAGGCTTACA 257
273 GCGCTCGGCAAGAGACTTGAAGTCAAGTCTTGTAGTGTGTGCTCACTTTGTGAAGT 332
258 GATGAAGGCAAGAGACTATGATGTTGTAAAACTTATGGAAGAGGTGCTTTGGTGAAGT 317
333 GCAGGTGTGAAGAGAAAAAGCAACGGGGACATCTATGCTATGAAAGTGATGAAGAGAA 392
318 GCAGTTGGTGTGTCACAAAGGATCCAGAAAGTTATGCTATGAAAGCTTCTTAGAAAGTT 377
393 GCGTTATTTGGCCGAGAGACAGTTTCAATTTTGTAGAAAGCGGACATATTTCTCG 452
378 TGAATATGATAAAAAGATCATGATTTCTGCTTTTGTGGAAAGAAAGATATTATGCGCTT 437
453 AAGCAAGACCCGCTGATCCGCCAATTAAGATGCTTTCAGACAAGAAATCAGCTTTA 512
438 TGCCATATGCCCCCTGGGTGTGCTTACCTTTTATGCTTTCAGATGATAGTATCTGTA 497
513 TCTGTCATGGAATATCAAGCTGAGGGGACTTGTCTCACTTTTGAATAGATATGAGGA 572
498 CATGCTAATGAGTACATGCTGAGGAGACCTTGTAAACCTTATGAGTAATTTATGATG- 556
573 CCAAGTATAGTAAACCTGATACAGTTTACCTAGTACAGTATTTGGCTGTCAAG 632
557 -----TGCCTGAAAAATGGCCAAATTTTACCTGCTGAGTGTCTTCTGCTGAGTGC 611
633 CGTTCATCTGATGGGATACGTGATCAGACATCAAGCTGAGAAACATTTCTGTGACCG 692
612 AATACACTCCATGGCTTTAATATCAGAGATGTGAAGCTGACAAACATGCTCTTGGATTA 671
693 CACAGGACATCAAGCTGTGTGTGATTTTGTGATCTCCCGGAAAGTAATCAAACAAGAT 752
672 ACATGACATCTAAATATGACAGATTTTGGACGTGTATGAAAGATGATGAACAGGCGAT 731
753 GGTGATGCAAAATCCGATTTGGAGCCCAAGATTAACATGCTCTGGAAGTCTGACTGT 812
732 GGTATATGTGATACAGAGCTTGAACACCGGATTTATATACCTTGAAGTTCTTAAATC 791
813 GATGAACGGGGATGAAAAAGCACCTACGGCTGACTGTGACTGTGTGATGAGGGGCT 872
792 ACAAGGGGGTGTAT-----GGTTTCTATGGCGGAAGATGTATGTGTGTGATGAGT 845
873 GATTCCTATGATGATGATTTATGAGAGATCCCCCTTGCAGAGGGAACCTCTGCCAAGC 932
846 TTTCTTTATGATGATGCTATGAGGGGATATCTCAATTTATGCGATTTCACTTGTAGAAC 905
933 CTTCATATACATTTATGAATTTCCAGCGGTTTGAATTTCCAGATGAACCCCAAGTGAG 992
906 ATATGCAAAATTTATGATCATAAAGATTTCACTGTGTTTCCCTGAAAGATGCAAAATTTT 965
993 CA 994
966 CA 967

RESULT 15
US-09-016-434-1105
; Sequence 1105, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREMITH
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1105:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4065 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEBANK
; CLONE: G1276900
; US-09-016-434-1105

Query Match 2.8%; Score 175.4; DB 4; Length 4065;
Best Local Similarity 51.1%; Pred. No. 2.1e-35;
Matches 469; Conservative 0; Mismatches 436; Indels 12; Gaps 2;

145 GAAGGATATATAGATCCCTCTTGTCTCTTTGAAGATGACATCAGCTCTCGATG 204
82 GATTTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 141
205 AAGATTAAGCAGTGAAGCACTTTGTCGGAAGTATCCGACACCATAGCTGAGTTACG 264
142 AAAAAAATAATGACACATTTTAAAGCAGATATAAGACACATATAATATAATACGA 201
265 GAGCTCCAGCTTGGCAAGAGACTTGAAGTCAAGATCTTGTGATGATGATGATGATGAT 324
202 GATTACCAATGAAGACGGAAGATATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 261
325 GGTGAAGTCAAGGTGTAAGAGAGAAAGCAACCGGGGACATCTATGATGAAGTGAAGT 384
262 GGAAGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 321
385 AAGAAAGAGCTTTATTTGGCCAGAGAGAGGATTTCAATTTTGTAGAAAGCGGAACATA 444
322 AGCAATTTGAATATGATTAAGAGATCTGATTTCTGTTTCTGGAAGAAAGGACATC 381
445 TTATTCGAAGACCAAGCCCGTGATCCCCCAATTAAGATATGCTTTCAGACAAAAAT 504
382 ATGCTTTTGGCAACAGCTTGGGTGTTGATGATTTTATGCAATTCACAAATGATGCT 441
505 CACCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 564
442 TATCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 501
565 TATGAGACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 624
502 TATGATG-----TGCCTGAAAAATGGGCAAGATCTATACGCAAGAGTATGTTCTTGA 555
625 GTTACACCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
556 TTGATGCAATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 615
685 GTTACCGCAGAGACATCAAGCTGTGATTTTGTGATCTGCGCGAAATGATTC 744

```
Db      616 CTGGATAAACTGGACATTTGAAGTATAGCAGATTTTGTACTGTATGAAGATGAATAAG 675
Qy      745 AACAAATGGTGATGCCAACTCCGATTTGGGACCCCAATTACATGGCTCTGAACTG 804
Db      676 GAAGGCATGGTACGATGTGTACAGCGGTTGGAAACCTGATTATATTTCCCTGAAGT- 734
Qy      805 CTGACTGTATGAACGGGATGGGAAAGCACCCTACGCGCTGAGCTGTGACTGTGATCA 864
Db      735 -----ATTAAATCCCAAGGTGATGGTTATTATGAAAGAAATGTGACTGTGATCG 789
Qy      865 GTGGGGGTGATTGCTATGAGATGATTTATGGGAGATCCCCCTTCGAGAGGGAACTCT 924
Db      790 GTTGGGGTATTATTATACGAAATGCTTGTAGGTGATACACCTTTTATGAGATTCTTG 849
Qy      925 GCCAGACCTTCAATACATTATGAATTCAGCGGTTTGAATTTCCAGATGACCC 984
Db      850 GTTGGAACTTACAGTAAATATGAAACATAAAATTCACCTTCTCTGATGATTAAT 909
Qy      985 AAAGTGAGCAGTACTTCTTGTATCTGATTCAAAGCTGTGTGGCGGCCAGAAAGAGAGA 1044
Db      910 GACATATCAAAAAGCAAAAAACCTTATTTGTGCTTCTTACTGACAGGGAAGTGAGG 969
Qy      1045 CTGAAGTTGAAGTCT 1061
Db      970 TTAGGGCGAAATGTGT 986
```

Search completed: November 15, 2003, 16:33:19
Job time : 276.099 secs

[illegible]

PF 20-DEC-2001; 2001MO-US50497.
XX
XX 27-DEC-2000; 2000US-258335P.
XX
XX (LEXI-) LEXICON GENETICS INC.
PA
XX
PI Yu X, Miranda M, Fiddle CJ;
XX
XX WPI; 2002-599796/64.
DR P-PSDB; ABB81927.
XX
XX Novel polynucleotide encoding human proteins that are structurally
PT similar to animal kinases, useful for drug screening, diagnosis, in
PT gene therapy of disorders and diseases e.g. cancer and pharmacogenomic
PT applications -
PS
PS Claim 1: Page 37-39; 50pp; English.
XX
XX The invention relates to a novel human protein that shares structural
CC similarity with animal kinases, including serine-threonine kinases,
CC particularly with animal rho-interacting kinases. The proteins of the
CC invention have nootropic and cyostatic activity. The polynucleotides may
CC have a use in gene therapy. The encoded novel polypeptides are useful for
CC generating antibodies, as reagents in diagnostic assays, for identifying
CC other cellular gene products related to NHP and as reagents in assays for
CC screening for compounds that are useful in the treatment of mental,
CC biological or medical disorders and diseases including cancer. The
CC sequence encodes a novel human kinase of the invention.
XX
XX Sequence 6165 BP; 1735 A; 1550 C; 1679 G; 1201 T; 0 other;
SQ
Query Match 100.0%; Score 6165; DB 24; Length 6165;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTGAAGTCAAAATATGAGCGCGGAAATCTTTGATGCTGCTGCTGAACCAT 60
DB 1 ATGTTGAAGTCAAAATATGAGCGCGGAAATCTTTGATGCTGCTGCTGAACCAT 60
QY 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCAA 120
DB 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCAA 120
QY 121 CACCAATGTCCTCTCTTCCCGAAGAGATATTAATGCTCTTGTCTTTGAA 180
DB 121 CACCAATGTCCTCTCTTCCCGAAGAGATATTAATGCTCTTGTCTTTGAA 180
QY 121 CACCAATGTCCTCTCTTCCCGAAGAGATATTAATGCTCTTGTCTTTGAA 180
DB 121 CACCAATGTCCTCTCTTCCCGAAGAGATATTAATGCTCTTGTCTTTGAA 180
QY 181 GAATGAGTCACTGCTCTGATGAAGATTAAGACCTGAGCACTTGTCCGAAGTAT 240
DB 181 GAATGAGTCACTGCTCTGATGAAGATTAAGACCTGAGCACTTGTGTCCGAAGTAT 240
QY 241 TCCGACACCATAGCTGATTAAGAGCTCCAGCTTCCGCAAGAGATTGAAATCAGA 300
DB 241 TCCGACACCATAGCTGATTAAGAGCTCCAGCTTCCGCAAGAGATTGAAATCAGA 300
QY 301 AGCTCTGAGGTTGTGCTCACTTGTGATGAGTGTGTAAGAGAAAGAACCGGG 360
DB 301 AGCTCTGAGGTTGTGCTCACTTGTGATGAGTGTGTAAGAGAAAGAACCGGG 360
QY 361 GACATCTATGCTATGAAGATGAAGAAAGAGCTTTATTTGGCCCGAGAGAGTTTCA 420
DB 361 GACATCTATGCTATGAAGATGAAGAAAGAGCTTTATTTGGCCCGAGAGAGTTTCA 420
QY 421 TTTTGTGAGGAAGAGGGAACATATTAATCTCGAAGCAAAAGCCGCTGATCCCAATTA 480
DB 421 TTTTGTGAGGAAGAGGGAACATATTAATCTCGAAGCAAAAGCCGCTGATCCCAATTA 480
QY 481 CAGTATGCTCTTGAAGCAAAATCACTTATCTGCTGATGAATATGCTGAGAGG 540
DB 481 CAGTATGCTCTTGAAGCAAAATCACTTATCTGCTGATGAATATGCTGAGAGG 540
QY 541 GACTCTGCTCACTTTGAAATGATATGAGACCAAGTTAGATAAACTGATACAGTTT 600
DB 541 GACTCTGCTCACTTTGAAATGATATGAGACCAAGTTAGATAAACTGATACAGTTT 600

DB 541 GACTCTGCTCACTTTGAAATGATATGAGACCAAGTTAGATAAACTGATACAGTTT 600
QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCCAGCGGTCATCTGATGGGATACGATCGA 660
DB 601 TACCTAGCTGAGCTGATTTTGGCTGTTCCAGCGGTCATCTGATGGGATACGATCGA 660
QY 661 GACATCAAGCTTGAGAACTTCTGTTGACCGACAGACATCAAGCTGTGTGATTTT 720
DB 661 GACATCAAGCTTGAGAACTTCTGTTGACCGACAGACATCAAGCTGTGTGATTTT 720
QY 721 GGAATCGCGCGGAAATGAAATTCAAACAGATGTAATGCCAACTCCGATTTGGGACC 780
DB 721 GGAATCGCGCGGAAATGAAATTCAAACAGATGTAATGCCAACTCCGATTTGGGACC 780
QY 781 CCAGATTACATGCTCTGAGAGTCTGATGATGAACGGGATGGAAGAGCACTTAC 840
DB 781 CCAGATTACATGCTCTGAGAGTCTGATGATGAACGGGATGGAAGAGCACTTAC 840
QY 841 GGCCTGAGCTGTGATCTGTGTGAGTGTGAGTGTGATGATGATGATTTATGGAGA 900
DB 841 GGCCTGAGCTGTGATCTGTGTGAGTGTGAGTGTGATGATGATGATTTATGGAGA 900
QY 901 TCCGCTTCCAGAGGGAACCTGTCAGAACCTTCAATTAACATTAATTTCCAGCG 960
DB 901 TCCGCTTCCAGAGGGAACCTGTCAGAACCTTCAATTAACATTAATTTCCAGCG 960
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAATCTTTGATCTGATTTCAAGC 1020
DB 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAATCTTTGATCTGATTTCAAGC 1020
QY 1021 TTGTTGTGCGGCAAGAAAGAGATGAACTTGAAGTCTTTGTGCTGCACTTTCTTC 1080
DB 1021 TTGTTGTGCGGCAAGAAAGAGATGAACTTGAAGTCTTTGTGCTGCACTTTCTTC 1080
QY 1081 TCTAAATGATCTGGAACAAATTCGTAATCTCTCCGCTGCTGCTCCACCTCAAG 1140
DB 1081 TCTAAATGATCTGGAACAAATTCGTAATCTCTCCGCTGCTGCTCCACCTCAAG 1140
QY 1141 TCTGAGATGACACTCCAAATTTTGTATGAACAGAGAAGATTCGTGGTTTCACTCT 1200
DB 1141 TCTGAGATGACACTCCAAATTTTGTATGAACAGAGAAGATTCGTGGTTTCACTCT 1200
QY 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTGCCGTTTGTGGGTTTTCG 1260
DB 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTGCCGTTTGTGGGTTTTCG 1260
QY 1261 TACAGCAAGCACTGAGGAAATCTTGTGATGATGAGTGTGCTGTGCTGAGTCTCC 1320
DB 1261 TACAGCAAGCACTGAGGAAATCTTGTGATGATGAGTGTGCTGTGCTGAGTCTCC 1320
QY 1321 CCGTCCAAAGCTAGCTCAATGGAAGAAATCTTCATCAAAAGCAAGAGCTTACAGAC 1380
DB 1321 CCGTCCAAAGCTAGCTCAATGGAAGAAATCTTCATCAAAAGCAAGAGCTTACAGAC 1380
QY 1381 TCTCAGAGCAAGTGTCAAGAGAGAGAGAGAAATGACCCGTTTCACTCGGAAGTCTCA 1440
DB 1381 TCTCAGAGCAAGTGTCAAGAGAGAGAGAGAAATGACCCGTTTCACTCGGAAGTCTCA 1440
QY 1441 GAGGTGAGGCTGTGCTTGTGATGAGAGAGGTTGAGGCTGAGAGCTGAGAGAG 1500
DB 1441 GAGGTGAGGCTGTGCTTGTGATGAGAGAGGTTGAGGCTGAGAGCTGAGAGAG 1500
QY 1501 TCCCTCTGAGAGAGACCTTGTCTACATCAATCAAGATGAGTGAAGCTTAAAGCGAAGT 1560
DB 1501 TCCCTCTGAGAGAGACCTTGTCTACATCAATCAAGATGAGTGAAGCTTAAAGCGAAGT 1560
QY 1561 TTGGAAGCAAGAGAGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620
DB 1561 TTGGAAGCAAGAGAGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620
QY 1621 GATATCAGAGAGAGAGCGGAACTCCAAAGAAATCAAAGAGAGAGATACAGGCTCAA 1680
DB 1621 GATATCAGAGAGAGAGCGGAACTCCAAAGAAATCAAAGAGAGAGATACAGGCTCAA 1680

Qy	1681	GTGGAGAAATGAGGTTGATGATGAAATCGATTGGAAGAGATCTTGTCTCAGCAAGAAGA	1740
Db	1681	GTGGAGAAATAGGGTTGATGATGAAATCGATTGGAAGAGATCTTGTCTCAGCAAGAAGA	1740
Qy	1741	CGGAGGATCTCTAGCAATCTAGCTGAGAGAGCTCCGCTTGTGCTGAAGAATTCAAG	1800
Db	1741	CGGAGGATCTCTAGCAATCTAGCTGAGAGAGCTCCGCTTGTGCTGAAGAATTCAAG	1800
Qy	1801	CGGAAAGCGACAGAAATGTCAGCATTAACGTGTTGAAGGCTTAAGGATCAAGGAAAGCTGGA	1860
Db	1801	CGGAAAGCGACAGAAATGTCAGCATTAACGTGTTGAAGGCTTAAGGATCAAGGAAAGCTGGA	1860
Qy	1861	GTGGAGGAATATGCCAATCTGGAAGAATCAAATGCTGAGCGACGCTCAAAATTCAGAG	1920
Db	1861	GTGGAGGAATATGCCAATCTGGAAGAATCAAATGCTGAGCGACGCTCAAAATTCAGAG	1920
Qy	1921	CTCCAAAGAGAAATCTGAGAAGGCTGTAAAGCGCAGCAAGAGGCGCACCGAGCTGCTCAG	1980
Db	1921	CTCCAAAGAGAAATCTGAGAAGGCTGTAAAGCGCAGCAAGAGGCGCACCGAGCTGCTCAG	1980
Qy	1981	AATATCCGCGCAAGAGAGCGAGCCGAGAGGAGCTTGAGAACTTGAGAACTGAGAACCGAG	2040
Db	1981	AATATCCGCGCAAGAGAGCGAGCCGAGAGGAGCTTGAGAACTTGAGAAACCGAG	2040
Qy	2041	GATTCCTCTGAAGGCGATCAGAAAGAGCTGTGGAAGCTGAGAAACGCGCGCATCTCTG	2100
Db	2041	GATTCCTCTGAAGGCGATCAGAAAGAGCTGTGGAAGCTGAGAAACGCGCGCATCTCTG	2100
Qy	2101	GAGAACAAAGTAAAGAGACTAGAGCCATGAGCGGTAGAGAAACAGACTGAAGGATGAC	2160
Db	2101	GAGAACAAAGTAAAGAGACTAGAGCCATGAGCGGTAGAGAAACAGACTGAAGGATGAC	2160
Qy	2161	ATCCGAGCAAAATCCCAACAGATCCAGAGATGCGTGAATTAATTCTGAGCTTGAAGAG	2220
Db	2161	ATCCGAGCAAAATCCCAACAGATCCAGAGATGCGTGAATTAATTCTGAGCTTGAAGAG	2220
Qy	2221	AAACCTCGGAGGCGCCCAAGCTCTCAGCCGAGCACTAAGAGCACTGAAAGCAACGAAAGAG	2280
Db	2221	AAACCTCGGAGGCGCCCAAGCTCTCAGCCGAGCACTAAGAGCACTGAAAGCAACGAAAGAG	2280
Qy	2281	CAGCATATGAGGAAAAAGATTAAAGTTTGAGCAATCAGATTAAGAAAGACTGCGCTGAC	2340
Db	2281	CAGCATATGAGGAAAAAGATTAAAGTTTGAGCAATCAGATTAAGAAAGACTGCGCTGAC	2340
Qy	2341	AAGGAGCACTGAGGAAACATGATGACAGAGACACGAGAGAGAGGCGCCATGAGAAAGGCGAA	2400
Db	2341	AAGGAGCACTGAGGAAACATGATGACAGAGAGACACGAGAGAGAGGCGCCATGAGAAAGGCGAA	2400
Qy	2401	ATTCTCAGCGAACGAAGGCGATGATCAATGCTATGATTTCCAAAGTCAGATCAGATCCCTGGA	2460
Db	2401	ATTCTCAGCGAACGAAGGCGATGATCAATGCTATGATTTCCAAAGTCAGATCAGATCCCTGGA	2460
Qy	2461	CAGAGGATTTGGAACCTGTCTGAAGCCAAATAAATTGACCAAAATAGCACTCTTTTACC	2520
Db	2461	CAGAGGATTTGGAACCTGTCTGAAGCCAAATAAATTGACCAAAATAGCACTCTTTTACC	2520
Qy	2521	CAAGGAAACATGAAGGCGCCAAGAAAGATGATTTCTGAACTCAGAGGCAACAGAAATTTTAC	2580
Db	2521	CAAGGAAACATGAAGGCGCCAAGAAAGATGATTTCTGAACTCAGAGGCAACAGAAATTTTAC	2580
Qy	2581	CTGGAGACACAGGCTGAGAAATTGAGAGGCCAGAACCGAAACTGAGAGAGGCACTGAGAG	2640
Db	2581	CTGGAGACACAGGCTGAGAAATTGAGAGGCCAGAACCGAAACTGAGAGAGGCACTGAGAG	2640
Qy	2641	AAGATTCAGCCACCAAGACCAAGTGAACAAATAATGCGTGTCTGGAATCTGAGACAGATTTG	2700
Db	2641	AAGATTCAGCCACCAAGACCAAGTGAACAAATAATGCGTGTCTGGAATCTGAGACAGATTTG	2700
Qy	2701	CGGAGGATCACTAGAGCAAGAGGAGCAAAACTGAGAGCTCAAGGCGCAGCTCAGAGAG	2760
Db	2701	CGGAGGATCACTAGAGCAAGAGGAGCAAAACTGAGAGCTCAAGGCGCAGCTCAGAGAG	2760

[illegible]

Db	3841	GTTCCTCTGCAATACATGAGCTGAAGCTGGCCCTTGGAAGAGAAAGCTCGCTGTGCA	3900
QY	3901	GAGCTAGAGAAAGCCCTTCAGAAAGACCCGATCGAGCTCCGGTCCGCCGGGAGGAACT	3960
Db	3901	GAGCTAGAGAAAGCCCTTCAGAAAGACCCGATCGAGCTCCGGTCCGCCGGGAGGAACT	3960
QY	3961	GCCCAACCGCAAGCAACGAGCCACCAACCCATCCAGCCGACCCGACGAGGACAG	4020
Db	3961	GCCCAACCGCAAGCAACGAGCCACCAACCCATCCAGCCGACCCGACGAGGACAG	4020
QY	4021	ATTCGCATGTCCGCCATGTGTGGGTCCGACAGACCCAGCCCATGTGCATGAGCTGTG	4080
Db	4021	ATTCGCATGTCCGCCATGTGTGGGTCCGACAGACCCAGCCCATGTGCATGAGCTGTG	4080
QY	4081	GCCCCGCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGGAATTTAGTGGCGCTT	4140
Db	4081	GCCCCGCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGGAATTTAGTGGCGCTT	4140
QY	4141	AAGAACGCATGCAACCAATATTTCTCAACCGATTCAACGTAGACTGAACATCGAGCC	4200
Db	4141	AAGAACGCATGCAACCAATATTTCTCAACCGATTCAACGTAGACTGAACATCGAGCC	4200
QY	4201	ACAAAGTGTGTGTGTGTGTGTGATTCGTGACCTTTGACCCGACGAGCTCAAAATGTCTC	4260
Db	4201	ACAAAGTGTGTGTGTGTGTGATTCGTGACCTTTGACCCGACGAGCTCAAAATGTCTC	4260
QY	4261	GAATGTCAAGTATGTCTACCCCAAGGTCTCAAGTCTCTTGACAGCACTGGGGCTTG	4320
Db	4261	GAATGTCAAGTATGTCTACCCCAAGGTCTCAAGTCTCTTGACAGCACTGGGGCTTG	4320
QY	4321	CCTGCTGAATATGCACACACTTCAACGAGGCGCTTGCCTGTGACAAATGATCTCCCA	4380
Db	4321	CCTGCTGAATATGCACACACTTCAACGAGGCGCTTGCCTGTGACAAATGATCTCCCA	4380
QY	4381	GGTCTTCAGACCAAGAGCCCAAGCAGCTTGCACTTGAAAGGTGATGAGGTGCC	4440
Db	4381	GGTCTTCAGACCAAGAGCCCAAGCAGCTTGCACTTGAAAGGTGATGAGGTGCC	4440
QY	4441	AGGAATTAACAAACGAGACAGAGGCTGGGACAGGAAGTACATTGTCTCGAGAGGATCA	4500
Db	4441	AGGAATTAACAAACGAGACAGAGGCTGGGACAGGAAGTACATTGTCTCGAGAGGATCA	4500
QY	4501	AAAGTCTCATTTATGACAAATGAAGCCAGAAAGCTGACAGAGCCGGTGAAGAAATTT	4560
Db	4501	AAAGTCTCATTTATGACAAATGAAGCCAGAAAGCTGACAGAGCCGGTGAAGAAATTT	4560
QY	4561	GAGCTGTGCCCTTCCGACGGGAGATGTATCTTTCATGTGTGCCGTTGTGCTCCGAATCTC	4620
Db	4561	GAGCTGTGCCCTTCCGACGGGAGATGTATCTTTCATGTGTGCCGTTGTGCTCCGAATCTC	4620
QY	4621	GCAAAATACAGCAAAAGCAGATGTCCCATATACATCTGAAGATGAAATCAACCCGCAACC	4680
Db	4621	GCAAAATACAGCAAAAGCAGATGTCCCATATACATCTGAAGATGAAATCAACCCGCAACC	4680
QY	4681	ACTGTGTGCCCGGAGAAACCTCTACTTGTCTAGACTCCAGCTTCCCTGCAAAACAGCCG	4740
Db	4681	ACTGTGTGCCCGGAGAAACCTCTACTTGTCTAGACTCCAGCTTCCCTGCAAAACAGCCG	4740
QY	4741	TGGGTCAACCGCTTGAATCAATGTGTGTGACAGGTGGAGAGTTTCTAGGAAAAAGCAGAA	4800
Db	4741	TGGGTCAACCGCTTGAATCAATGTGTGTGACAGGTGGAGAGTTTCTAGGAAAAAGCAGAA	4800
QY	4801	GCTATGTCTAACTGCTTGAATCTCCTGCTGAACTGAAAGGTGATGACCCGCTTAGAC	4860
Db	4801	GCTATGTCTAACTGCTTGAATCTCCTGCTGAACTGAAAGGTGATGACCCGCTTAGAC	4860
QY	4861	ATGAATCGACAGCTGCCCTTCAAGTACCAAGTGTGTGTGTGAGCCACCGAGAAAGGCTC	4920
Db	4861	ATGAATCGACAGCTGCCCTTCAAGTACCAAGTGTGTGTGTGAGCCACCGAGAAAGGCTC	4920
QY	4921	TACCCCTGAAATGTCTTGAATACTCCCTAACCATGTCCAGGAATTGAGACATGCTTC	4980

Db	4921	TACGCCCTGAATTGTCCTTGAAAAA	CTCCCTAACCCATGCTCCAGGAATTGAGAGATCTTC	4980
QY	4981	CAAAATTATATATATCAAGACCTGAGAGATCACTCATGTAGAGAGAAAGACGGGCA		5040
Db	4981	CAAAATTATATATATCAAGACCTGAGAGATCACTCATGTAGAGAGAAAGACGGGCA		5040
QY	5041	CTGTGTCTTGTGACGTGTAAAGAAAGTAAACAGTCCCTGGCCAGTCCCACTGACCTGCC		5100
Db	5041	CTGTGTCTTGTGACGTGTAAAGAAAGTAAACAGTCCCTGGCCAGTCCCACTGACCTGCC		5100
QY	5101	CAGCCCGACATCTCACCCACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA		5160
Db	5101	CAGCCCGACATCTCACCCACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA		5160
QY	5161	GGCAAGATTGAGAACGGGCTTGATCTGTGACACCATGCCACAAAGTCGTATTCTC		5220
Db	5161	GGCAAGATTGAGAACGGGCTTGATCTGTGACACCATGCCACAAAGTCGTATTCTC		5220
QY	5221	CGCTACACAAAACTCAGCAAAATCTGCATCCGAAAGAGATAGACCTCAGACCC		5280
Db	5221	CGCTACACAAAACTCAGCAAAATCTGCATCCGAAAGAGATAGACCTCAGACCC		5280
QY	5281	TGCAGCTGTATCACTTCAACCAATTACAGTATCTCATTTGAAACCAATAATTCTAGAA		5340
Db	5281	TGCAGCTGTATCACTTCAACCAATTACAGTATCTCATTTGAAACCAATAATTCTAGAA		5340
QY	5341	ATCGACATGAAGCAGTACACGCTGTGAGAAATCTGTGATTAAGATATGACATTCCTTGGCA		5400
Db	5341	ATCGACATGAAGCAGTACACGCTGTGAGAAATCTGTGATTAAGATATGACATTCCTTGGCA		5400
QY	5401	CTGTGTGTTTGGCCGCTCTTCCACAGCTCCCTGTCAATCGACAGGTGAACAGC		5460
Db	5401	CTGTGTGTTTGGCCGCTCTTCCACAGCTCCCTGTCTTCAATCGACAGGTGAACAGC		5460
QY	5461	GCAGGGCAGCAGAGAGATCTTGCTGTGTTTCCACGAATTTGAGAGTTCGTGATCT		5520
Db	5461	GCAGGGCAGCAGAGAGATCTTGCTGTGTTTCCACGAATTTGAGAGTTCGTGATCT		5520
QY	5521	TACGGAAGCCTTACCCGCAACAGATCTCAAGTGAAGTGCCTTACCTTGGCTTGGCC		5580
Db	5521	TACGGAAGCCTTACCCGCAACAGATCTCAAGTGAAGTGCCTTACCTTGGCTTGGCC		5580
QY	5581	TACAGAGAACCCTATCTGTTTGTGACCCCATCTCAACCACTCCGAAAGTAATTGAGATCCAG		5640
Db	5581	TACAGAGAACCCTATCTGTTTGTGACCCCATCTCAACCACTCCGAAAGTAATTGAGATCCAG		5640
QY	5641	GCACGCTCTCAGCAGGGAACCCCTGCCAGCGTACTGTGACATCCCGAAACCCGCGCTAC		5700
Db	5641	GCACGCTCTCAGCAGGGAACCCCTGCCAGCGTACTGTGACATCCCGAAACCCGCGCTAC		5700
QY	5701	CTGGGCCCCCTGCATTTTCTCAGAGAGCATTTACTTGGCGTCTCTCATACAGGATTAATTA		5760
Db	5701	CTGGGCCCCCTGCATTTTCTCAGAGAGCATTTACTTGGCGTCTCTCATACAGGATTAATTA		5760
QY	5761	AGGGCTATTTTGCTCAAGGAAAACTGTGAAGAGTCGGGCACTGAACCAACCCGCGGCG		5820
Db	5761	AGGGCTATTTTGCTCAAGGAAAACTGTGAAGAGTCGGGCACTGAACCAACCCGCGGCG		5820
QY	5821	CCGTCCAACCTCCGACAGACCCCAACAAGCGAGCCACCCACGTTACAAGAGACATC		5880
Db	5821	CCGTCCAACCTCCGACAGACCCCAACAAGCGAGCCACCCACGTTACAAGAGACATC		5880
QY	5881	ACCAAGCGGTGGCTTCACACCCAGGCGCCCGAAAGGCCCCGACCCCGCAGAGGCCA		5940
Db	5881	ACCAAGCGGTGGCTTCACACCCAGGCGCCCGAAAGGCCCCGACCCCGCAGAGGCCA		5940
QY	5941	AGCACACCCACCGCTACCGCGAGGGGCGAGCTCGAAGTCTCTTGCC		6000
Db	5941	AGCACACCCACCGCTACCGCGAGGGGCGAGCTCGAAGTCTCTTGCC		6000
QY	6001	CGCCCCCTGAGAGGAGAGATCCCGGCGGATCTCAGACCGCGAGAGAGCGTCC		6060
Db	6001	CGCCCCCTGAGAGGAGAGATCCCGGCGGGAATCTCAGACCGCGAGAGAGCGTCC		6060

Qy 6061 CCCGAGAGGCTTTTGAAGACAGCAGGAGGCGGCTGCTGCGGAGCCGTGAGACC 6120
 Db 6061 CCCGAGAGGCTTTTGAAGACAGCAGGAGGCGGCTGCTGCGGAGCCGTGAGACC 6120
 Qy 6121 CCGCTGTCGAGGTGAACAGAGTCTGGAGCCAGTCTTCAATATA 6165
 Db 6121 CCGCTGTCGAGGTGAACAGAGTCTGGAGCCAGTCTTCAATATA 6165

RESULT 2

AD38864
 ID AAD38864 standard; cDNA; 6298 BP.

AC AAD38864;

DT 23-SEP-2002 (first entry)

DE Human kinase (PKIN) -21 cDNA.

Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;
 acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;
 asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;
 development; hepatitis; cardiovascular; hypertension; drug screening;
 myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;
 fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;
 hypercholesterolaemia; obesity; gene therapy; cytosolic; anti-hiv;
 neuroprotective; hepatotropic; hypotensive; cardiac; nephrotropic;
 hyperlipidaemia; enzyme; gene; se.

OS Homo sapiens.

Key: Location/Qualifiers
 FT 55..6219
 CDS /*tag= a
 /product= "Human kinase (PKIN) -21"

W0200233099-A2.

25-APR-2002.

20-OCT-2001; 2001WO-US47728.

20-OCT-2000; 2000US-242410P.
 27-OCT-2000; 2000US-244068P.
 03-NOV-2000; 2000US-245708P.
 09-NOV-2000; 2000US-247672P.
 16-NOV-2000; 2000US-249565P.
 22-NOV-2000; 2000US-252730P.
 01-DEC-2000; 2000US-250807P.

(INCY-) INCYTE GENOMICS INC.

PI Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvinu C;
 PI Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;
 PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribolety CM;
 PI Lai PG, Reardon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;
 PI Thangavelu K, Khan FA, Ison CH;

WPI; 2002-454603/48.

P-PSDB; AAE24150.

New human kinase polypeptide, for diagnosing, preventing and treating
 cancer, immune system disorders, growth and development disorders,
 cardiovascular disorders and lipid disorders

Claim 5; Page 207-209; 210pp; English.

The invention relates human kinases (PKIN) and their corresponding
 nucleic acid sequences. PKIN and its DNA are useful for diagnosing,
 treating and preventing cancer, an immune system disorder (e.g.,
 acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,
 asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders

CC affecting growth and development (e.g., arteriosclerosis, cirrhosis,
 CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial
 CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty
 CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolemia,
 CC hyperlipidaemia, obesity), and for assessing the effects of exogenous
 CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a
 CC condition or a disease associated with the expression of PKIN in a
 CC biological sample. A composition comprising PKIN or an agonist or
 CC antagonist of PKIN is useful for treating a disease or condition
 CC associated with decreased or increased expression of functional PKIN.
 CC PKIN is useful in a number of drug screening techniques and to analyse
 CC the proteome of a tissue or cell type. PKIN DNA is useful for creating
 CC knockin humanised animals or transgenic animals to model human diseases,
 CC and in somatic or germline gene therapy. The present sequence is human
 CC PKIN cDNA.

Sequence 6298 BP; 1772 A; 1585 C; 1720 G; 1221 T; 0 other;

Query Match 99.9%; Score 6161.8; DB 24; Length 6298;

Best Local Similarity 100.0%; Pred. No. 0; Matches 6163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGTTGAAGTTCAATATGAGCGGGAATCCTTGTGATGCTGCTGAACCCATT 60
 Db 55 ATGTTGAAGTTCAATATGAGCGGGAATCCTTGTGATGCTGCTGAACCCATT 114
 Qy 61 GCCAGCCGAGCTTCAGAGCTGAATCTGTTCTTCCAGGGGAAACACCTTTATGACTCA 120
 Db 115 GCCAGCCGAGCTTCAGAGCTGAATCTGTTCTTCCAGGGGAAACACCTTTATGACTCA 174
 Qy 121 CAGCAGATGTCCTCTCTTCCCGAAGAGATTAATGATGCTCTTGTCTTTGAA 180
 Db 175 CAGCAGATGTCCTCTCTTCCCGAAGAGATTAATGATGCTCTTGTCTTTGAA 234
 Qy 181 GAATGCAGTCAGCTGCTGATGAATTAAGACGTGAGCACTTGTCCGGAGAT 240
 Db 235 GAATGCAGTCAGCTGCTGATGAATTAAGACGTGAGCACTTGTCCGGAGAT 294
 Qy 241 TCCGACACCATAGCTGATTAAGAGAGCTTCAGGCTTCGGAAGAGCTTGAAGTCA 300
 Db 295 TCCGACACCATAGCTGATTAAGAGAGCTTCAGGCTTCGGAAGAGCTTGAAGTCA 354
 Qy 301 AGCTTGTAGTGTGTGCTCTTCTGAAAGTGCAGGTGTGAAGAGAAACACCGGG 360
 Db 355 AGCTTGTAGTGTGTGCTCTTCTGAAAGTGCAGGTGTGAAGAGAAACACCGGG 414
 Qy 361 GACATCATGCTATGAAGATGAAGAGAGAGCTTTATGGCCGAGGAGAGTTCA 420
 Db 415 GACATCATGCTATGAAGATGAAGAGAGAGCTTTATGGCCGAGGAGAGTTCA 474
 Qy 421 TTTTGAAGAGAGGGAACATATATCTCGAAGACAAAGCCGTGATCCCAATTA 480
 Db 475 TTTTGAAGAGAGGGAACATATATCTCGAAGACAAAGCCGTGATCCCAATTA 534
 Qy 481 CAGTATGCTTTAGAGCAAAATTCACCTTATCTGCTCATGGAATATCAGCTTGAGGG 540
 Db 535 CAGTATGCTTTAGAGCAAAATTCACCTTATCTGCTCATGGAATATCAGCTTGAGGG 594
 Qy 541 GACTTGTGCTACTTTGAATAGATATGAGGACCATGTTGATGAATGAACCTGATACGTT 600
 Db 595 GACTTGTGCTACTTTGAATAGATATGAGGACCATGTTGATGAATGAACCTGATACGTT 654
 Qy 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGAGCTTCATCTGATGAGATACGATCGA 660
 Db 655 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGAGCTTCATCTGATGAGATACGATCGA 714
 Qy 661 GACATGAAGCTGAGACATCTCCGTTGACCGGACGACACATCAAGCTGTGATTTT 720
 Db 715 GACATGAAGCTGAGACATCTCCGTTGACCGGACGACACATCAAGCTGTGATTTT 774
 Qy 721 GGATGCGCGCAAAATGATTCAAACAAGATGTTGAATGCCAACTCCCGATTTGGAGCC 780
 Db 775 GGATGCGCGCAAAATGATTCAAACAAGATGTTGAATGCCAACTCCCGATTTGGAGCC 834

```
QY 781 CCAGATTACATGCTCTCTGAAGTCTGACTGTGATGAAGGGGATGAAAAAGGCACCTAC 840
Db 835 CCAGATTACATGCTCTCTGAAGTCTGACTGTGATGAAGGGGATGAAAAAGGCACCTAC 894
QY 841 GGCCTGACCTGTGACTGTGTGTCAGTGGGCGTGAATGCTATGAGATGATTTATGGGAGA 900
Db 895 GGCCTGACCTGTGACTGTGTGTCAGTGGGCGTGAATGCTATGAGATGATTTATGGGAGA 954
QY 901 TCCCTCTGCGAAGGAACTCTGCGAAGCTTCCATATACATTAATGATTTCCAGCGG 960
Db 955 TCCCTCTGCGAAGGAACTCTGCGAAGCTTCCATATACATTAATGATTTCCAGCGG 1014
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTTGATCTGAATCAAGC 1020
Db 1015 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTTGATCTGAATCAAGC 1074
QY 1021 TTGTTGTGGCGCCAGAAAGAGACTGAAGTTTGAAGTCTTGTGCGCATCTTCTTC 1080
Db 1075 TTGTTGTGGCGCCAGAAAGAGACTGAAGTTTGAAGTCTTGTGCGCATCTTCTTC 1134
QY 1081 TCTAAATTTGACTGGAACAACATTCGTAACTCTCTCCCTCCCTGTTCCACCTCAAG 1140
Db 1135 TCTAAATTTGACTGGAACAACATTCGTAACTCTCTCCCTCCCTGTTCCACCTCAAG 1194
QY 1141 TCTGAGATGACACCTCCAAATTTTGAATGAACGAGAGAAATTCGTGGATTCACTCT 1200
Db 1195 TCTGAGATGACACCTCCAAATTTTGAATGAACGAGAGAAATTCGTGGATTCACTCT 1254
QY 1201 CCGTCCAGCTGAGCCCTCAGGCTTCTCGSGTGAAGAACTGCCGTTTGTGGGTTTTCG 1260
Db 1255 CCGTCCAGCTGAGCCCTCAGGCTTCTCGSGTGAAGAACTGCCGTTTGTGGGTTTTCG 1314
QY 1261 TACAGCAAGCACTGGGGAATTTCTGTAGATCTGAATCTGTGTGCGGCTCGACTCC 1320
Db 1315 TACAGCAAGCACTGGGGAATTTCTGTAGATCTGAATCTGTGTGCGGCTCGACTCC 1374
QY 1321 CCTGCAAGACTGATCCATGGAAGAACTTCTCATCAAGAAAGCAAGAGCTTAACAAGC 1380
Db 1375 CCTGCAAGACTGATCCATGGAAGAACTTCTCATCAAGAAAGCAAGAGCTTAACAAGC 1434
QY 1381 TCTCAGACAAGTGTCAAGAATGAGAGAGAAATGACCCGGTTCATCGAGAGTGTCA 1440
Db 1435 TCTCAGACAAGTGTCAAGAATGAGAGAGAAATGACCCGGTTCATCGAGAGTGTCA 1494
QY 1441 GAGGTGGAGCTGTCTTAAGTCAAGAGAGGTGAGCTGAAAGCTTGAAGCTCAGAGA 1500
Db 1495 GAGGTGGAGCTGTCTTAAGTCAAGAGAGGTGAGCTGAAAGCTTGAAGCTCAGAGA 1554
QY 1501 TCCCTCTGCGAAGGAACTTGTCTACCTACATCAAGATGCAAGATGCTTAAGGAGAT 1560
Db 1555 TCCCTCTGCGAAGGAACTTGTCTACCTACATCAAGATGCAAGATGCTTAAGGAGAT 1614
QY 1561 TTGAGCAAGCAAGGATGAGGTGTCCAGAGAGATGAACAAGCACTGCAAGCTTCCAT 1620
Db 1615 TTGAGCAAGCAAGGATGAGGTGTCCAGAGAGATGAACAAGCACTGCAAGCTTCCAT 1674
QY 1621 GATATCAGAGAGCAGAGCCGGAAGTCCAAAGAAATCAAGAGCAGAGTACCAAGCTCAA 1680
Db 1675 GATATCAGAGAGCAGAGCCGGAAGTCCAAAGAAATCAAGAGCAGAGTACCAAGCTCAA 1734
QY 1681 GTGGAAGAAATGAGGTTGATGATGATCAGTTTGAAGAGATCTTGTCTCAGCAAGAA 1740
Db 1735 GTGGAAGAAATGAGGTTGATGATGATCAGTTTGAAGAGATCTTGTCTCAGCAAGAA 1794
QY 1741 CGAGATGATCTCTAAGATCTGAGCTGAGAGTCTCGGCTTGTCTGAAGAAATTCAG 1800
Db 1795 CGAGATGATCTCTAAGATCTGAGCTGAGAGTCTCGGCTTGTCTGAAGAAATTCAG 1854
QY 1801 CGGAAGGAGAGATGTCAACATTAATCTGTTGAAGCTTAAGATCAAGGAGAGCTGAA 1860
Db 1855 CGGAAGGAGAGATGTCAACATTAATCTGTTGAAGCTTAAGATCAAGGAGAGCTGAA 1914
QY 1861 GTGGAGAAATATGCGAACTGAGAGAAATCAATGTCTAGACAGAGCTCAAAATTCAGAG 1920
Db 1915 GTGGAGAAATATGCGAACTGAGAGAAATCAATGTCTAGACAGAGCTCAAAATTCAGAG 1974
QY 1921 CTCCAAAGAAATCTGAGAGAGCTGTAAAGCCAGACGAGAGCCACGAGCTCTCGAG 1980
Db 1975 CTCCAAAGAAATCTGAGAGAGCTGTAAAGCCAGACGAGAGCCACGAGCTCTCGAG 2034
QY 1981 AATATCCGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 2035 AATATCCGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
QY 2041 GATTTCTTGAAGGATCAAGAAAGCTGTGGAAGCTGAAGAAAGCCGCAATTCCTG 2100
Db 2095 GATTTCTTGAAGGATCAAGAAAGAGCTGTGGAAGCTGAAGAAAGCCGCAATTCCTG 2154
QY 2101 GAGAACAAAGTAAAGAGCTAGAGACATGAGCGGTAGAGAAACAGACTGAAGATGAC 2160
Db 2155 GAGAACAAAGTAAAGAGCTAGAGACATGAGCGGTAGAGAAACAGACTGAAGATGAC 2214
QY 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGAGTGAATTAATCTGAGCTCGAAGAG 2220
Db 2215 ATCCAGACAAATCCCAACAGATCCAGAGATGAGTGAATTAATCTGAGCTCGAAGAG 2274
QY 2221 AAACATCGGAGGCCCAAGTCTGAGCCCAAGCACTGAAGTCACTGAAAGAGAGAG 2280
Db 2275 AAACATCGGAGGCCCAAGTCTGAGCCCAAGCACTGAAGTCACTGAAAGAGAGAG 2334
QY 2281 CAGCACTATGAGGAAAGATTAATGTTGACATTCAGATTAAGAAAGACTGTGAC 2340
Db 2335 CAGCACTATGAGGAAAGATTAATGTTGACATTCAGATTAAGAAAGACTGTGAC 2394
QY 2341 AAGAGACACTGAGAGACATGATGAGAGACACGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2395 AAGAGACACTGAGAGACATGATGAGAGACACGAGAGAGAGAGAGAGAGAGAGAGAG 2454
QY 2401 ATTCTCAGCGAACAGAGAGCCATGATCAATGCTATGATTCAGATCAGATCCTGAA 2460
Db 2455 ATTCTCAGCGAACAGAGAGCCATGATCAATGCTATGATTCAGATCAGATCCTGAA 2514
QY 2461 CAGAGGATTTGAGAACTGCTGGAAGCCAAATTAATCTGAGCAAAATAGCACTTTTAC 2520
Db 2515 CAGAGGATTTGAGAACTGCTGGAAGCCAAATTAATCTGAGCAAAATAGCACTTTTAC 2574
QY 2521 CAAAGGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGCTCAGGCAACAGAAATTTAC 2580
Db 2575 CAAAGGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGCTCAGGCAACAGAAATTTAC 2634
QY 2581 CTGAGACACAGGCTGGGAAAGTTGAGGCCCCAGAAACGAGAACTGAGAGAGCTGAG 2640
Db 2635 CTGAGACACAGGCTGGGAAAGTTGAGGCCCCAGAAACGAGAACTGAGAGAGCTGAG 2694
QY 2641 AAGATCAGCCCAAGACCAAGTGAACAAGATGAGGCTGAGAACTGAGAGCAAGATTG 2700
Db 2695 AAGATCAGCCCAAGACCAAGTGAACAAGATGAGGCTGAGAACTGAGAGCAAGATTG 2754
QY 2701 CGGAGGCTCAGTCTAGAGACGAGAGAGAGAACTGAGGCTCAAGCGCCAGCTCAGAG 2760
Db 2755 CGGAGGCTCAGTCTAGAGACGAGAGAGAGAACTGAGGCTCAAGCGCCAGCTCAGAG 2814
QY 2761 CTACAGCTCTCTCTGAGAGAGCGAGTCAAGTTGACAGGCTCTGAGGCTGACAGGCG 2820
Db 2815 CTACAGCTCTCTCTGAGAGAGCGAGTCAAGTTGACAGGCTCTGAGGCTGACAGGCG 2874
QY 2821 GCCCTGAGAGCCAGCTTCCGCAAGGCAAGCAAGCTGGAAGAGACCAAGCGAGAGCT 2880
Db 2875 GCCCTGAGAGCCAGCTTCCGCAAGGCAAGCAAGCTGGAAGAGACCAAGCGAGAGCT 2934
QY 2881 GAAGAGAGATCCAGGCACTCAAGGCAATAGAGATGAATCCAGCGCAATTTGATGCT 2940
Db 2935 GAAGAGAGATCCAGGCACTCAAGGCAATAGAGATGAATCCAGCGCAATTTGATGCT 2994
QY 2941 CTTGTAAACAGCTGTACTATATCAGACCTGAGAGAGAGAGCTAAACCAAGTCAAG 3000
```

Db 2995 CTTGTAACAGCTGTAAGTATCAAGACCTGGAGGAGCAGCTTAACAGCTGACCGAG 3054
Qy 3001 GACAACGCTGAATCTCAACACCAAACTTCTAATTGTCCAAACAACTGATGAGCTTCT 3060
Db 3055 GACAACGCTGAATCTCAACACCAAACTTCTAATTGTCTCAAACTGATGAGCTTCT 3114
Qy 3061 GCGCGCAACGACGAGATTGTATCAACTGGAGTGAAGTGAACCTCTCCGCGGAGATC 3120
Db 3115 GCGCGCAACGACGAGATTGTATCAACTGGAGTGAAGTGAACCTCTCCGCGGAGATC 3174
Qy 3121 ACGGACGAGAGATGACGCTTACGACGAGAGAAACGATGAGGCTCTGAAGACACG 3180
Db 3175 ACGGACGAGAGATGACGCTTACGACGAGAGAAACGATGAGGCTCTGAAGACACG 3234
Qy 3181 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGCTTAAACGATGAGCTTGA 3240
Db 3235 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGCTTAAACGATGAGCTTGA 3294
Qy 3241 AAGAGAGGAGAGATGAGAGGCTGAGAGAGCTCTGAGTGAAGAAATCCCACTTTGAG 3300
Db 3295 AAGAGAGGAGAGATGAGAGGCTGAGAGAGCTCTGAGTGAAGAAATCCCACTTTGAG 3354
Qy 3301 TGTGAGGTTGAGAGCTGAGAGAGATGCTGAGACCCGAGAAACAGAGAGGCGAGAGCC 3360
Db 3355 TGTGAGGTTGAGAGCTGAGAGAGATGCTGAGACCCGAGAAACAGAGAGGCGAGAGCC 3414
Qy 3361 GATCAGGAGATCAACCCGATCTCCGACAGTGTGAGAGCTGCGATGGAAGAGACAGAGCT 3420
Db 3415 GATCAGGAGATCAACCCGATCTCCGACAGTGTGAGAGCTGCGATGGAAGAGACAGAGCT 3474
Qy 3421 GAGATTTCTGCTCTGACGACAGGCTCTCAAGAGACGAAAGCTGAAGGCGGAGAGCTTCT 3480
Db 3475 GAGATTTCTGCTCTGACGACAGGCTCTCAAGAGACGAAAGCTGAAGGCGGAGAGCTTCT 3534
Qy 3481 GACAAGCTCAATGACCTGAGAAAGACATGCTATGCTTGAATGAATGACCCGAGAGCTT 3540
Db 3535 GACAAGCTCAATGACCTGAGAAAGACATGCTATGCTTGAATGAATGACCCGAGAGCTT 3594
Qy 3541 CACGACGAGCTGAGAGATGAAACGAGAGCTCAACAGAGCTTCTGGAAGAGCAAGCCAA 3600
Db 3595 CACGACGAGCTGAGAGATGAAACGAGAGCTCAACAGAGCTTCTGGAAGAGCAAGCCAA 3654
Qy 3601 TTACAGAGCAGATGAGACCTGAGAAATCAATTTCCGCTGATCAAGAGCTTGA 3660
Db 3655 TTACAGAGCAGATGAGACCTGAGAAATCAATTTCCGCTGATCAAGAGCTTGA 3714
Qy 3661 GAAAGCTTAAGATGAGGCTGATCTATGAGAGACAGAAAGAGTGAAGTGAAGCTG 3720
Db 3715 GAAAGCTTAAGATGAGGCTGATCTATGAGAGACAGAAAGAGTGAAGTGAAGCTG 3774
Qy 3721 GAAAACATTCAGGTTCTCTATTTCTCATGAAAAGTGAAGAGGCACTATTTCTCA 3780
Db 3775 GAAAACATTCAGGTTCTCTATTTCTCATGAAAAGTGAAGAGGCACTATTTCTCA 3834
Qy 3781 CAAACCAAACTCATGATTTTCTGCAAGCCAAATGAGACCACTGCTTAAAGAAAAG 3840
Db 3835 CAAACCAAACTCATGATTTTCTGCAAGCCAAATGAGACCACTGCTTAAAGAAAAG 3894
Qy 3841 GTTCTCTGACGATACATGAGCTGAAAGCTGACCTGAGAGAGAGAAAGCTGCTGCA 3900
Db 3895 GTTCTCTGACGATACATGAGCTGAAAGCTGACCTGAGAGAGAGAAAGCTGCTGCA 3954
Qy 3901 GAGCTAGAGAGAGCCCTTTCAGAAAGCCGCTATGAGAGCTGCGCCCGGAGAGAGCT 3960
Db 3955 GAGCTAGAGAGAGCCCTTTCAGAAAGCCGCTATGAGAGCTGCGCCCGGAGAGAGCT 4014
Qy 3961 GCCACCGCAAGAGCAACGAGACCAACCCATCAACCGCGGAGAGAGAGAGAGAGAG 4020
Db 4015 GCCACCGCAAGAGCAACGAGACCAACCCATCAACCGCGGAGAGAGAGAGAGAGAG 4074
Qy 4021 ATGCGCATGTCGCGCATGCTGCGGTGCGAGAGACCAAGCTCCATGAGCCTGCTG 4080

Db 4075 ATGCGCATGTCGCGCATGCTGCGGTGCGCAAGAGCAACGCCAAGTCCATGAGCTGCTG 4134
Qy 4081 GCCCGCCATCCAGCCGAGAAAGAGTCTTCAATCCAGAGAAATTTAGTCGAGCTT 4140
Db 4135 GCCCGCCATCCAGCCGAGAAAGAGTCTTCAATCCAGAGAAATTTAGTCGAGCTT 4194
Qy 4141 AAGGAGCAGATGACCAATATTTCTTCAACGATTTCAACGATGAGACTGAACATGAGAGC 4200
Db 4195 AAGGAGCAGATGACCAATATTTCTTCAACGATTTCAACGATGAGACTGAACATGAGAGC 4254
Qy 4201 ACAAAGT 4260
Db 4255 ACAAAGT 4314
Qy 4261 GAATGTGAGT 4320
Db 4315 GAATGTGAGT 4374
Qy 4321 CTGCTGAATATGTCACACACTTTCACCGAGGCTTCTGCGTGAACAAATGAACTCCCA 4380
Db 4375 CTGCTGAATATGTCACACACTTTCACCGAGGCTTCTGCGTGAACAAATGAACTCCCA 4434
Qy 4381 GGTCTCGACGCAAGAGAGCCGAGAGAGCTTGTGACCTGTGAAGAGGTGATGAAGTGTGCC 4440
Db 4435 GGTCTCGACGCAAGAGAGCCGAGAGAGCTTGTGACCTGTGAAGAGGTGATGAAGTGTGCC 4494
Qy 4441 AGGATTAACAAACGAGAGACAGAGCTGGGACAGAAATGATCAATGCTCTGAGAGATCA 4500
Db 4495 AGGATTAACAAACGAGAGACAGAGCTGGGACAGAAATGATCAATGCTCTGAGAGATCA 4554
Qy 4501 AAGTCTCTCATTTATGCAATTAAGCCAGAGAAAGCTGACAGAGCCGATGAAATTT 4560
Db 4555 AAGTCTCTCATTTATGCAATTAAGCCAGAGAAAGCTGACAGAGCCGATGAAATTT 4614
Qy 4561 GAGCTGTGCTTCCGAGCGGAGATGATCTATTTCAATGAGTGTGCTGTGCTTCCGAACTC 4620
Db 4615 GAGCTGTGCTTCCGAGCGGAGATGATCTATTTCAATGAGTGTGCTGTGCTTCCGAACTC 4674
Qy 4621 GCAAAATACAGCCAAAGAGATGTCCCATACATCTGAAGATGGAATCTCACCCGACACCC 4680
Db 4675 GCAAAATACAGCCAAAGAGATGTCCCATACATCTGAAGATGGAATCTCACCCGACACCC 4734
Qy 4681 ACCTGTGCGCGGAGAAACCTCTACTTGTGCTGAGCTTCCGAGCTTCCGTAACACAGCCG 4740
Db 4735 ACCTGTGCGCGGAGAAACCTCTACTTGTGCTGAGCTTCCGAGCTTCCGTAACACAGCCG 4794
Qy 4741 TGGGTACCGCCTTAATGATGATGTGTGCGAGGTGCGAGAGTTCTTGAAGAAAAACAGAA 4800
Db 4795 TGGGTACCGCCTTAATGATGATGTGTGCGAGGTGCGAGAGTTCTTGAAGAAAAACAGAA 4854
Qy 4801 GCTGATGCTAACTGCTTGAAGAACTCCCTGCTGAACCTGGAAGGTGATGACCCGCTAGAC 4860
Db 4855 GCTGATGCTAACTGCTTGAAGAACTCCCTGCTGAACCTGGAAGGTGATGACCCGCTAGAC 4914
Qy 4861 ATGAATGCAAGCTGCTCTTCACTGAGCAAGAGTGTGTGAGGAGCAACGAGAGAGGCTC 4920
Db 4915 ATGAATGCAAGCTGCTCTTCACTGAGCAAGAGTGTGTGAGGAGCAACGAGAGAGGCTC 4974
Qy 4921 TAGCGCTGAATGTCTTGAAGAACTCCCTTAACCAATGCTCCAGAGATTTGAGAGATCTTC 4980
Db 4975 TAGCGCTGAATGTCTTGAAGAACTCCCTTAACCAATGCTCCAGAGATTTGAGAGATCTTC 5034
Qy 4981 CAAATTTATATTAATCAAGAGCTGAGAGATCACTATGATAGTACAGAGAGAGAGAGAGAG 5040
Db 5035 CAAATTTATATTAATCAAGAGCTGAGAGATCACTATGATAGTACAGAGAGAGAGAGAGAG 5094
Qy 5041 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5100
Db 5095 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5154
Qy 5101 CAGCCCGACATCAACCCCAATTTTGAAGCTGTCAAGAGGCTGCACTTGTGTGGGCA 5160
Db 5155 CAGCCCGACATCAACCCCAATTTTGAAGCTGTCAAGAGGCTGCACTTGTGTGGGCA 5214

QY 5161 GGCAGATGAGAACGGGCTCTGCACTCTGAGCCATGCCAGCAAAAGTCATCTC 5220
 DB 5215 GGCAGATGAGAACGGGCTCTGCACTCTGAGCCATGCCAGCAAAAGTCATCTC 5274
 QY 5221 CGCTACAGCAAAACCTCAGCAATATCTGCATCCGAAAAGATGAGACTCAGAGCCC 5280
 DB 5275 CGCTACAGCAAAACCTCAGCAATATCTGCATCCGAAAAGATGAGACTCAGAGCCC 5334
 QY 5281 TGGAGCTGATCCACTTCAACCAATACATGATCTCATTTGGAACCAATAAATCTACGAA 5340
 DB 5335 TGGAGCTGATCCACTTCAACCAATACATGATCTCATTTGGAACCAATAAATCTACGAA 5394
 QY 5341 ATGACATGAGAGAGTACAGCTCGAGAAATTCCTGATTAAGATGACATTCCTTGGCA 5400
 DB 5395 ATGACATGAGAGAGTACAGCTCGAGAAATTCCTGATTAAGATGACATTCCTTGGCA 5454
 QY 5401 CCTGCTGTGTTTGGCCGCTCTTCCAGAGCTTCCCTGTCTCATCTGTCAGGTGAGACAGC 5460
 DB 5455 CCTGCTGTGTTTGGCCGCTCTTCCAGAGCTTCCCTGTCTCATCTGTCAGGTGAGACAGC 5514
 QY 5461 GCAGGGCAGAGAGAGTACTGCTGTGTTTCCAGCAATTTGGAGTGTGTCGATGATCT 5520
 DB 5515 GCAGGGCAGAGAGAGTACTGCTGTGTTTCCAGCAATTTGGAGTGTGTCGATGATCT 5574
 QY 5521 TACGAAAGAGTACGCGCAGACAGATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTTGCC 5580
 DB 5575 TACGAAAGAGTACGCGCAGACAGATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTTGCC 5634
 QY 5581 TACGAAAGAGTACGCTGTTTGTGACCCACTTCAACTCAAGTAATTTGAGATTCAG 5640
 DB 5635 TACGAAAGAGTACGCTGTTTGTGACCCACTTCAACTCAAGTAATTTGAGATTCAG 5694
 QY 5641 GCAGGCTCTCAGAGAGAGCCCTGCGCAGCGTACTGAGCAATCCCGAACCAGCGCTAC 5700
 DB 5695 GCAGGCTCTCAGAGAGAGCCCTGCGCAGCGTACTGAGCAATCCCGAACCAGCGCTAC 5754
 QY 5701 CTGGGCTCTGCACTTCTTCAAGAGCATTTACTTGGCGTCTTCATACAGATAAATTA 5760
 DB 5755 CTGGGCTCTGCACTTCTTCAAGAGCATTTACTTGGCGTCTTCATACAGATAAATTA 5814
 QY 5761 AGGATCATTTGCTGCAAGAGGAACCTGTGAAGAGTCGGGCACTGAACACCCGGGGC 5820
 DB 5815 AGGATCATTTGCTGCAAGAGGAACCTGTGAAGAGTCGGGCACTGAACACCCGGGGC 5874
 QY 5821 CCGTCCACTCCCGCAGCAGCCCAACAAGCAGAGCCCAACGTAACAAGACATC 5880
 DB 5875 CCGTCCACTCCCGCAGCAGCCCAACAAGCAGAGCCCAACGTAACAAGACATC 5934
 QY 5881 ACCAAGCGGTGCTTCCAGCCAGCGCGCCGAGAGGCCCAAGCCCGCAGAGCCA 5940
 DB 5935 ACCAAGCGGTGCTTCCAGCCAGCGCGCCGAGAGGCCCAAGCCCGCAGAGCCA 5994
 QY 5941 AGCAACCCCAACCGCTACCGCGAGAGGGCGGACCGAGTGTGCGAGGCAAGTCTCCGTC 6000
 DB 5995 AGCAACCCCAACCGCTACCGCGAGAGGGCGGACCGAGTGTGCGAGGCAAGTCTCCGTC 6054
 QY 6001 CGCCCTCTGAGAGAGAGAGTCCCGCGCGAGTACTCAAGCAGCGAGAGAGCGGTCC 6060
 DB 6055 CGCCCTCTGAGAGAGAGAGTCCCGCGCGAGTACTCAAGCAGCGAGAGAGCGGTCC 6114
 QY 6061 CCGCGAGAGCTGTTTGAAGCAGAGAGAGGGCGGCTGCTCGCGGAGAGCCGTGAGAGAC 6120
 DB 6115 CCGCGAGAGCTGTTTGAAGCAGAGAGAGGGCGGCTGCTCGCGGAGAGCCGTGAGAGAC 6174
 QY 6121 CCGCTGTCCAGAGTGAACAAGGTCTGGAGACAGTCTTCAATATA 6165
 DB 6175 CCGCTGTCCAGAGTGAACAAGGTCTGGAGACAGTCTTCAATATA 6219

RESULT 3
 AAL55214
 ID AAL55214 standard; DNA: 6165 BP.

XX AAL55214;
 AC 01-MAY-2003 (first entry)
 XX
 DE Human CRK encoding DNA sequence, SEQ ID No 1.
 XX
 KW Anorectic; hypotensive; cardiast; antilipemic; cerebroprotective;
 KW Anticough; osteoporotic; antirheumatic; cytosolic; antidepressant;
 KW Immunomodulator; antihuman; tranquiliser; antiparkinsonian; nootropic;
 KW Neuroprotective; antineurotoxic; antidiabetic; analgesic;
 KW human citron rho/rac-interacting kinase; enzyme; CRK; ameliorating;
 KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;
 KW central nervous system disorder; chronic obstructive pulmonary disease;
 KW diabetes; pain; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..6165
 FT /*tag= a
 FT /product= "Human CRK protein"
 XX
 EN W02003004523-A1.
 XX
 PD 16-JAN-2003.
 XX
 PF 28-JUN-2002; 2002MO-EP07156.
 XX
 PR 02-JUL-2001; 2001US-301841P.
 PR 11-DEC-2001; 2001US-338651P.
 PR 25-APR-2002; 2002US-375014P.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Zhu Z;
 XX
 DR WPI; 2003-221576/21.
 DR P-PSDB; AAO26959.
 XX
 PT New human citron rho/rac-interacting kinase (CRK) polypeptide and
 PT polynucleotide, useful in preventing, ameliorating or treating diseases
 PT associated with human CRK dysfunction, e.g. obesity, diabetes or
 PT Alzheimer's disease -
 XX
 PS Example 1; Fig 1; 237pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding a human
 CC citron rho/rac-interacting kinase polypeptide. The isolated
 CC polynucleotide comprises a 6165 or 8603 base pair sequence, given in the
 CC specification. The human citron rho/rac-interacting kinase (CRK)
 CC polypeptide and polynucleotide are useful in preventing, ameliorating, or
 CC treating diseases associated with human CRK dysfunction such as obesity
 CC and obesity-associated comorbidities (e.g. hypertension, coronary artery
 CC disease, hyperlipidaemia, stroke, gout, osteoarthritis, some types of
 CC cancer including endometrial, breast, prostate and colon cancer),
 CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood
 CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's
 CC disease), chronic obstructive pulmonary disease or diabetes. These can
 CC also be used to treat pain associated with the disorders. The human CRK
 CC polypeptide is also useful in diagnostic assays or in genetic testing.
 CC The expression vector or the reagent is useful in preparing a medicament
 CC for modulating the activity of a human CRK in a disease, e.g. obesity, a
 CC central nervous system disorder, or chronic obstructive pulmonary
 CC disease. The fusion protein is useful for generating antibodies against a
 CC CRK polypeptide and for use in various assay systems. The methods are
 CC useful in producing and detecting the polynucleotide and polypeptide and
 CC in screening for agents that modulate the activity of the human CRK
 CC polypeptide. This polynucleotide sequence represents a DNA sequence
 CC encoding a human CRK protein of the invention.
 CC
 SQ Sequence 6165 BP; 1735 A; 1549 C; 1680 G; 1201 T; 0 other;

Query Match 99.9%; Score 6160.2; DB 25; Length 6165;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60
 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60
 61 GCCAGCGGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACACACCTTATGATCA 120
 61 GCCAGCGGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACACACCTTATGATCA 120
 121 CAGCAGATGCTCTCTCTTTCCCGAAGAGGATATTAGATGCCCTTTGTTCTTGA 180
 121 CAGCAGATGCTCTCTCTTTCCCGAAGAGGATATTAGATGCCCTTTGTTCTTGA 180
 181 GAATGCACTCAGCTGCTCTGATGATGATGATGATGATGATGATGATGATGAT 240
 181 GAATGCACTCAGCTGCTCTGATGATGATGATGATGATGATGATGATGATGAT 240
 241 TCCGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 241 TCCGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 301 AGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 301 AGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 361 GACATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 361 GACATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 421 TTTTGTGAGAGAGCGGAACATATATCTCGAAGCAAAAGCCGTGATCCCAATTA 480
 421 TTTTGTGAGAGAGCGGAACATATATCTCGAAGCAAAAGCCGTGATCCCAATTA 480
 481 CAGATGCTTCTGAGCAAAATATCATCTTATGCTGATGATGATGATGATGATGAT 540
 481 CAGATGCTTCTGAGCAAAATATCATCTTATGCTGATGATGATGATGATGATGAT 540
 541 GACTTGTGCTCACTTTTGAATATGATGATGATGATGATGATGATGATGATGAT 600
 541 GACTTGTGCTCACTTTTGAATATGATGATGATGATGATGATGATGATGATGAT 600
 601 TACTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 601 TACTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGACATCAAGCTGATGATTTT 720
 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGACATCAAGCTGATGATTTT 720
 721 GGATCTGCGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 721 GGATCTGCGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 780
 781 CCAGATTACATGCTCTCGGAGTCTGATGATGATGATGATGATGATGATGATGAT 840
 781 CCAGATTACATGCTCTCGGAGTCTGATGATGATGATGATGATGATGATGATGAT 840
 841 GGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 841 GGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 901 TCCCTCTTCCAGAGGAGGAGCTGCGAGAACTTCAATTAATGATTTTCCAGCG 960
 901 TCCCTCTTCCAGAGGAGGAGCTGCGAGAACTTCAATTAATGATTTTCCAGCG 960
 961 TTTTGAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 961 TTTTGAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 1021 TTTTGAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

1021 TTTTGTGCGCGCAAAAGAGAGCTGAAGTTTGAAGTCTTTGCTGCACTCTTCTTC 1080
 1081 TCTTAAATTTGATGAGCAAACTTTGTAATCTCTCTCTCTCTCTCTCTCTCTCAAG 1140
 1081 TCTTAAATTTGATGAGCAAACTTTGTAATCTCTCTCTCTCTCTCTCTCTCTCAAG 1140
 1141 TCTGAGATGACACCTCCCAATTTTATGAAACAGAGAAATGCTGAGTTTCACTCT 1200
 1141 TCTGAGATGACACCTCCCAATTTTATGAAACAGAGAAATGCTGAGTTTCACTCT 1200
 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGAGTGAAGAACTGCGGTTTGTGAGGTTTTCG 1260
 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGAGTGAAGAACTGCGGTTTGTGAGGTTTTCG 1260
 1261 TACAGCAAGGCACTGAGGATTTTGTGATGATGATGATGATGATGATGATGATGAT 1320
 1261 TACAGCAAGGCACTGAGGATTTTGTGATGATGATGATGATGATGATGATGATGAT 1320
 1321 CCTGCAAGACTAGCTCCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAC 1380
 1321 CCTGCAAGACTAGCTCCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAC 1380
 1381 TCTCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 1381 TCTCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 1441 GAGTGTGAGGCTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 1441 GAGTGTGAGGCTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 1501 TCCCTTCTGAGAGCAGGACCTTCTCATCATCAAGAAATGAGTAAAGCGAAAT 1560
 1501 TCCCTTCTGAGAGCAGGACCTTCTCATCATCAAGAAATGAGTAAAGCGAAAT 1560
 1561 TTGAGAGCAAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 1561 TTGAGAGCAAGCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 1621 GATATCAGAGCAGAGCGGAGCTCCAGAAATCAAGAGAGATGATGATGATGATGAT 1680
 1621 GATATCAGAGCAGAGCGGAGCTCCAGAAATCAAGAGAGATGATGATGATGATGAT 1680
 1681 GTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1681 GTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1741 CGAGATGATCTTACGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 1741 CGAGATGATCTTACGAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 1801 CGAGAGCGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 1801 CGAGAGCGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 1861 GTGGAGAAATATGAGAACTGAGAAATGATGATGATGATGATGATGATGATGATGAT 1920
 1861 GTGGAGAAATATGAGAACTGAGAAATGATGATGATGATGATGATGATGATGATGAT 1920
 1921 CTCGAGAGAACTGAGAAAGCTGTAAAGCCAGACAGAGCCAGAGCTGTGAG 1980
 1921 CTCGAGAGAACTGAGAAAGCTGTAAAGCCAGACAGAGCCAGAGCTGTGAG 1980
 1981 AATATCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
 1981 AATATCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
 2041 GATTCCTTCTGAGAGCATGAGAAAGCTGTGAGAGCTGAGAGAGCGCGCATTCCTG 2100
 2041 GATTCCTTCTGAGAGCATGAGAAAGCTGTGAGAGCTGAGAGAGCGCGCATTCCTG 2100
 2101 GAGAACAGGTAAGAGCTAGAGCATGAGAGCTAGAGAAACAGACTGAGAGTAC 2160

Db 2101 GAGAACAAAGGTAAGAGACTAGAGACCATGAGCCGTAGAGAAAACAGACTGAAAGATGAC 2160
Qy 2161 ATTCGACAAAATCCCAACAGATCCAGAGATGCTGATAAATTTCTGAGCTCGAAGAG 2220
Db 2161 ATTCGACAAAATCCCAACAGATCCAGAGATGCTGATAAATTTCTGAGCTCGAAGAG 2220
Qy 2221 AAAACATCGGAGAGCCCAAGCTCTCAGCCCGACCTTGAAGTGCACCTGAAAACAGAAAGAG 2280
Db 2221 AAAACATCGGAGAGCCCAAGCTCTCAGCCCGACCTTGAAGTGCACCTGAAAACAGAAAGAG 2280
Qy 2281 CAGCACTATGAGAGAAAAGATTAAAGTTGGAATGCAATTCAGATTAAGAAAGAACTGCTGAC 2340
Db 2281 CAGCACTATGAGAGAAAAGATTAAAGTTGGAATGCAATTCAGATTAAGAAAGAACTGCTGAC 2340
Qy 2341 AAGAGACACTGAGAGAACATGATGACAGACACAGAGAGAGAGCCCATGAGAGAGGCAAA 2400
Db 2341 AAGAGACACTGAGAGAACATGATGACAGACACAGAGAGAGAGCCCATGAGAGAGGCAAA 2400
Qy 2401 ATTCTCAGCGAAACAGAGGCGATGATCAATGCTATGAGATTCGAAGATCAGATCCCTGAA 2460
Db 2401 ATTCTCAGCGAAACAGAGGCGATGATCAATGCTATGAGATTCGAAGATCAGATCCCTGAA 2460
Qy 2461 CAGAGGATTTGGAACCTGCTGAAAGCCAAATAACTTGCAGCAATATGACAGCTTTTAAAC 2520
Db 2461 CAGAGGATTTGGAACCTGCTGAAAGCCAAATAACTTGCAGCAATATGACAGCTTTTAAAC 2520
Qy 2521 CAAAGGAACTGAGAGGCGCCAGAGAGAGATGATTTCTGAACTCGAGCAAACAGAAATTTTAC 2580
Db 2521 CAAAGGAACTGAGAGGCGCCAGAGAGAGATGATTTCTGAACTCGAGCAAACAGAAATTTTAC 2580
Qy 2581 CTGAGAGACAGAGGCGTGGAGAGTTGGAAGCCCAAGAACCCAGAACTGAGAGAGAGCTGAG 2640
Db 2581 CTGAGAGACAGAGGCGTGGAGAGTTGGAAGCCCAAGAACCCAGAACTGAGAGAGAGCTGAG 2640
Qy 2641 AAGATCAAGCCACAGAGACCAAGTGAACAAGATCGGCTGCTGAACTGGAACAAGATTG 2700
Db 2641 AAGATCAAGCCACAGAGACCAAGTGAACAAGATCGGCTGCTGAACTGGAACAAGATTG 2700
Qy 2701 CGGAGAGTCACTGAGAGACAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAG 2760
Db 2701 CGGAGAGTCACTGAGAGACAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAG 2760
Qy 2761 CTACAGCTCTCCCTGAGAGAGGCGAGTCAAGTTGAGAGCCCTGAGAGAGCTGACAGGCG 2820
Db 2761 CTACAGCTCTCCCTGAGAGAGGCGAGTCAAGTTGAGAGCCCTGAGAGAGCTGACAGGCG 2820
Qy 2821 GCCCTGAGAGCCAGCTTCCGACAGCGAAGACAGAGCTGAAAGAGACCAAGCAGAGAGCT 2880
Db 2821 GCCCTGAGAGCCAGCTTCCGACAGCGAAGACAGAGCTGAAAGAGACCAAGCAGAGAGCT 2880
Qy 2881 GAAAGAGAGATCCAGGACCTCAGGACATGAGATGAATTCAGCGCAAAATTTGATGCT 2940
Db 2881 GAAAGAGAGATCCAGGACCTCAGGACATGAGATGAATTCAGCGCAAAATTTGATGCT 2940
Qy 2941 CTTGTAACAGCTGTACTGTATCAGAGACTGAGAGAGAGCTAAACACACTACAGAG 3000
Db 2941 CTTGTAACAGCTGTACTGTATCAGAGACTGAGAGAGAGCTAAACACACTACAGAG 3000
Qy 3001 GACAAAGCTGAACTCAACCAAACTTCTACTGTTCCAAACAACTCGATGAGAGCTTCT 3060
Db 3001 GACAAAGCTGAACTCAACCAAACTTCTACTGTTCCAAACAACTCGATGAGAGCTTCT 3060
Qy 3061 GGGCCCAACGACGAGATTGTACAACTCGAAGTGAATGACATCTCCCGCGGAGATC 3120
Db 3061 GGGCCCAACGACGAGATTGTACAACTCGAAGTGAATGACATCTCCCGCGGAGATC 3120
Qy 3121 ACGGAGAGAGATGACAGCTTACAGCCAGAGCAAAAGATGAGAGCTGAAAGACAGAG 3180
Db 3121 ACGGAGAGAGATGACAGCTTACAGCCAGAGCAAAAGATGAGAGCTGAAAGACAGAG 3180
Qy 3181 TGCAACATGCTGAGAGAAAGCTATGATTTGAGAGCCCTTAAACGATGAGCTGCTGAA 3240
Db 3181 TGCAACATGCTGAGAGAAAGCTATGATTTGAGAGCCCTTAAACGATGAGCTGCTGAA 3240

Qy 3241 AAAGACGCGAGTGGAGAGGCTTGAGAGAGCGTCTGAGTATGATGAATATCCAGTTTGAAG 3300
Db 3241 AAAGACGCGAGTGGAGAGGCTTGAGAGAGCGTCTGAGTATGATGAATATCCAGTTTGAAG 3300
Qy 3301 TGTGGGTTTGAAGCTGCGAGAGATGCTGACACCGAGAAAACAGAGCAGAGGCGAGAGCC 3360
Db 3301 TGTGGGTTTGAAGCTGCGAGAGATGCTGACACCGAGAAAACAGAGCAGAGGCGAGAGCC 3360
Qy 3361 GATCAGCGGATACCGAGCTTCTGCAAGTGTGTGAGCTGAGCTGAAAGAGACCAAGGCT 3420
Db 3361 GATCAGCGGATACCGAGCTTCTGCAAGTGTGTGAGCTGAGCTGAAAGAGACCAAGGCT 3420
Qy 3421 GAGATTTCTGCTCTGAGAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGAGGCTCTCT 3480
Db 3421 GAGATTTCTGCTCTGAGAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGAGGCTCTCT 3480
Qy 3481 GACAAAGCTCAATGACCTGAGAGAGAGATGATGCTTGAATGAATGATGATGATGATGAT 3540
Db 3481 GACAAAGCTCAATGACCTGAGAGAGAGATGATGCTTGAATGAATGATGATGATGATGAT 3540
Qy 3541 CAGCAGAGCTGAGAGATGAAACGAGAGCTCAAAACAGAGCTTCTGAAAGACCAAGCCAAA 3600
Db 3541 CAGCAGAGCTGAGAGATGAAACGAGAGCTCAAAACAGAGCTTCTGAAAGACCAAGCCAAA 3600
Qy 3601 TTACAGCAGCAGATGAGACCTGACAGAAATACATTTTCCGCTGACTCAAGACTGCA 3660
Db 3601 TTACAGCAGCAGATGAGACCTGACAGAAATACATTTTCCGCTGACTCAAGACTGCA 3660
Qy 3661 GAAAGCTTGAATCGGAGCTGATCTAAGAGACAGAAAGAGTGAAGCTTGAAGTGAAGCTG 3720
Db 3661 GAAAGCTTGAATCGGAGCTGATCTAAGAGACAGAAAGAGTGAAGCTTGAAGTGAAGCTG 3720
Qy 3721 GAAAACATTCAGGTTCTCTATTTCTATGTAAGAAAGTGAAGAAAGGAGCACTATTTCTCAA 3780
Db 3721 GAAAACATTCAGGTTCTCTATTTCTATGTAAGAAAGTGAAGAAAGGAGCACTATTTCTCAA 3780
Qy 3781 CAAACCAAACTCATTTGATTTTCTGAAAGCCAAAATGACCAACCTGCTAAAAAGAAAAG 3840
Db 3781 CAAACCAAACTCATTTGATTTTCTGAAAGCCAAAATGACCAACCTGCTAAAAAGAAAAG 3840
Qy 3841 GTTCTCTGCAATCAATGAGCTGAGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 3841 GTTCTCTGCAATCAATGAGCTGAGAGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Qy 3901 GAGCTAAG 3960
Db 3901 GAGCTAAG 3960
Qy 3961 GCCCAGCGCAAG 4020
Db 3961 GCCCAGCGCAAG 4020
Qy 4021 ATGCGCATGTCGCGCATGTCGAGTCCGACAGAGCAAGCCAGTGCATGAGCTGCTG 4080
Db 4021 ATGCGCATGTCGCGCATGTCGAGTCCGACAGAGCAAGCCAGTGCATGAGCTGCTG 4080
Qy 4081 GCGCCGCGCATGCGCGAG 4140
Db 4081 GCGCCGCGCATGCGCGAG 4140
Qy 4141 AAGGAG 4200
Db 4141 AAGGAG 4200
Qy 4201 ACAAAGTGTGCTGT 4260
Db 4201 ACAAAGTGTGCTGT 4260
Qy 4261 GAAATGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4320
Db 4261 GAAATGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4320

QY 4321 CCTGTGATATATGACACACTTCAACGAGGCTTCTGCGGTGACAAATGAACTCCCA 4380
Db 4321 CCTGTGATATATGACACACTTCAACGAGGCTTCTGCGGTGACAAATGAACTCCCA 4380
QY 4381 GGTCTCCAGACCAAGAGCCAGACGAGCTTGGACCTGGAAGGTGATGAAGTGGCC 4440
Db 4381 GGTCTCCAGACCAAGAGCCAGACGAGCTTGGACCTGGAAGGTGATGAAGTGGCC 4440
QY 4441 AGGAATTAACAAAGAGACACCAAGGCTGGAGACAGAAAGTCAATTTCTGAGAGATCA 4500
Db 4441 AGGAATTAACAAAGAGACACCAAGGCTGGAGACAGAAAGTCAATTTCTGAGAGATCA 4500
QY 4501 AAATCTCATTTTATGACATGAGAGCCAGAGAGTGGAGACGAGGCGGTGGAAGATTT 4560
Db 4501 AAATCTCATTTTATGACATGAGAGCCAGAGAGTGGAGACGAGGCGGTGGAAGATTT 4560
QY 4561 GAGCTGTGCTTCCGACGAGGATGTATCTAATTCATGAGTGGCTGTGCTTCCGAACTC 4620
Db 4561 GAGCTGTGCTTCCGACGAGGATGTATCTAATTCATGAGTGGCTGTGCTTCCGAACTC 4620
QY 4621 GCAAAATACAGCAAGAGAGATGTCCTATACATCTGAAAGATGGAATCTCACCCGACACC 4680
Db 4621 GCAAAATACAGCAAGAGAGATGTCCTATACATCTGAAAGATGGAATCTCACCCGACACC 4680
QY 4681 ACCTGTGCGCCGAGAGAACCTCTACTGTGCTAGCTCCAGCTTCCCTGACAAACAGGCG 4740
Db 4681 ACCTGTGCGCCGAGAGAACCTCTACTGTGCTAGCTCCAGCTTCCCTGACAAACAGGCG 4740
QY 4741 TGGGTACCGCCTTAAATAGTTGTGACAGGTGGAGAGTTCTGAGGAAAAAGCAAA 4800
Db 4741 TGGGTACCGCCTTAAATAGTTGTGACAGGTGGAGAGTTCTGAGGAAAAAGCAAA 4800
QY 4801 GGTGATGCTAACTGCTTGGAGAACTCCCTGCTGAAACTGGAAGGTGATGACCGTCTAGAC 4860
Db 4801 GGTGATGCTAACTGCTTGGAGAACTCCCTGCTGAAACTGGAAGGTGATGACCGTCTAGAC 4860
QY 4861 ATGAATGACCGCTGCTTCACTGACACAGTGGTGTGAGGAGCAACGAGAGAGGCTC 4920
Db 4861 ATGAATGACCGCTGCTTCACTGACACAGTGGTGTGAGGAGCAACGAGAGAGGCTC 4920
QY 4921 TACGCTTGAATGCTTGAAGAACTCCCTAACCAGATGCTCCAGAAATTGAGAGATCTTC 4980
Db 4921 TACGCTTGAATGCTTGAAGAACTCCCTAACCAGATGCTCCAGAAATTGAGAGATCTTC 4980
QY 4981 CAAATTTATATTATCAAGAGCTGAGAGAGCTACTATGATGACAGAGAGAGCGGCA 5040
Db 4981 CAAATTTATATTATCAAGAGCTGAGAGAGCTACTATGATGACAGAGAGAGCGGCA 5040
QY 5041 CTGTGTCTGTGACGTAAGAGAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5100
Db 5041 CTGTGTCTGTGACGTAAGAGAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5100
QY 5101 CAGCCCGACATCTCAACCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160
Db 5101 CAGCCCGACATCTCAACCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160
QY 5161 GGCAGATTTGAGAGAGGAGCTGACATCTGAGCCATGCGCAAGCTGCACTTGTGTGGGCA 5220
Db 5161 GGCAGATTTGAGAGAGGAGCTGACATCTGAGCCATGCGCAAGCTGCACTTGTGTGGGCA 5220
QY 5221 CGCTACACGAAACCTCTGACAAATCTGATCCGAAAGATAGAGCTTCAAGGCC 5280
Db 5221 CGCTACACGAAACCTCTGACAAATCTGATCCGAAAGATAGAGCTTCAAGGCC 5280
QY 5281 TGCAGCTGTATCCATTTCAACATTTAGATGCTGATGGAACCAATTAATTTCAAGAA 5340
Db 5281 TGCAGCTGTATCCATTTCAACATTTAGATGCTGATGGAACCAATTAATTTCAAGAA 5340
QY 5341 ATCGACATGAGACAGTACCGCTCGAGAGATTCCTGGAATGAACCAATTCCTTGGCA 5400
Db 5341 ATCGACATGAGACAGTACCGCTCGAGAGATTCCTGGAATGAACCAATTCCTTGGCA 5400
QY 5401 CCTGTGTGTTTGGCGCTCTTCCAAAGCTTCCCTGTCTCAATGCTGACAGGTGAACAGC 5460

Db 5401 CCTGTGTGTTTGGCGCTCTTCCAAAGCTTCCCTGTCTCAATGCTGACAGGTGAACAGC 5460
QY 5461 GCAGGAGAGAGAGAGATCTGTGTGTTTCAACGAATTTGAGTGTGTGATTTCT 5520
Db 5461 GCAGGAGAGAGAGAGATCTGTGTGTTTCAACGAATTTGAGTGTGTGATTTCT 5520
QY 5521 TACGGAAGACGTACCCGACAGAGATCTCAAGTGAAGTGTGCTTACCTTTGAGCTTGGC 5580
Db 5521 TACGGAAGACGTACCCGACAGAGATCTCAAGTGAAGTGTGCTTACCTTTGAGCTTGGC 5580
QY 5581 TACGGAAGACGTATCTGTGTGTTTGAACCACTTCACTCACTGGAATTAATGATTCAG 5640
Db 5581 TACGGAAGACGTATCTGTGTGTTTGAACCACTTCACTCACTGGAATTAATGATTCAG 5640
QY 5641 GCAGCTCTGAG 5700
Db 5641 GCAGCTCTGAG 5700
QY 5701 CTGAGGCTGTGCAATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5760
Db 5701 CTGAGGCTGTGCAATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5760
QY 5761 AGGCTCATTTGTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5820
Db 5761 AGGCTCATTTGTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5820
QY 5821 CCGTCCACTTCCGACAG 5880
Db 5821 CCGTCCACTTCCGACAG 5880
QY 5881 ACCAG 5940
Db 5881 ACCAG 5940
QY 5941 AGCAGACATCCAGCTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
Db 5941 AGCAGACATCCAGCTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
QY 6001 CGCCCTTGAAG 6060
Db 6001 CGCCCTTGAAG 6060
QY 6061 CCGGAGAGAGCTGTTTGAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6120
Db 6061 CCGGAGAGAGCTGTTTGAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6120
QY 6121 CCGCTGTCCAGAGTGAACAGAGTCTGAGACAGTCTTCAATATA 6165
Db 6121 CCGCTGTCCAGAGTGAACAGAGTCTGAGACAGTCTTCAATATA 6165
RESULT 4
AAL55215
ID AAL55215 standard; DNA; 8603 BP.
XX AAL55215;
XX 01-MAY-2003 (first entry)
DE Human CR1K related DNA sequence, SEQ ID No 4.
XX
XX Anorectic; hypotensive; cardiast; antilipemic; cerebroprotective;
KW antiob; osteopathic; antiarthritic; cytoprotective; antidepressant;
KW immunomodulator; antineoplastic; tranquilizer; antiparkinsonian; neurotropic;
KW neuroprotective; antineoplastic; antidiabetic; analgesic;
KW human cation rho/rac-interacting kinase; enzyme; CR1K; ameliorating;
KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;
KW central nervous system disorder; chronic obstructive pulmonary disease;
KW diabetes; pain; de.
XX
XX Homo sapiens.
OS

PN W02003004523-AL.
 XX
 PD 16-JAN-2003.
 XX
 PF 28-JUN-2002; 2002WO-EP07156.
 XX
 PR 02-JUL-2001; 2001US-301841P.
 PR 11-DEC-2001; 2001US-338651P.
 PR 25-APR-2002; 2002US-375014P.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Zhu Z;
 DR WPI; 2003-221576/21.
 XX
 PT New human citron rho/rac-interacting kinase (CRIK) polypeptide and
 PT polynucleotide, useful in preventing, ameliorating or treating diseases
 PT associated with human CRIK dysfunction, e.g. obesity, diabetes or
 PT Alzheimer's disease
 XX
 PS Disclosure; Fig 4; 237p; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding a human
 CC citron rho/rac-interacting kinase polypeptide. The isolated
 CC polynucleotide comprises a 6165 or 8603 base pair sequence, given in the
 CC specification. The human citron rho/rac-interacting kinase (CRIK)
 CC polypeptide and polynucleotide are useful in preventing, ameliorating, or
 CC treating diseases associated with human CRIK dysfunction, such as obesity
 CC and obesity-associated comorbidities (e.g. hypertension, coronary artery
 CC disease, hyperlipidemia, stroke, gout, osteoarthritis, some types of
 CC cancer including endometrial, breast, prostate and colon cancer),
 CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood
 CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's
 CC disease), chronic obstructive pulmonary disease, or diabetes. These can
 CC also be used to treat pain associated with the disorders. The human CRIK
 CC polypeptide is also useful in diagnostic assays or in genetic testing.
 CC The expression vector or the reagent is useful in preparing a medicament
 CC for modulating the activity of a human CRIK in a disease, e.g. obesity, a
 CC central nervous system disorder, or chronic obstructive pulmonary
 CC disease. The fusion protein is useful for generating antibodies against a
 CC CRIK polypeptide and for use in various assay systems. The methods are
 CC useful in producing and detecting the polynucleotide and polypeptide and
 CC in screening for agents that modulate the activity of the human CRIK
 CC polypeptide. This polynucleotide sequence represents a DNA sequence
 CC relating to the human CRIK protein of the invention.
 XX
 SQ Sequence 8603 BP; 2305 A; 2206 C; 2215 G; 1877 T; 0 other;
 Query Match 99.9%; Score 6160.2; DB 25; Length 8603;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 301 AGCTCTGTAGCTGTGTGCTCACTTTGCTGAAGTCGAGTGTGTAAGAGAGAAACACCGG 360
 Db 301 AGCTCTGTAGCTGTGTGCTCACTTTGCTGAAGTCGAGTGTGTAAGAGAGAAACACCGG 360
 QY 361 GACATCATGCTATGAAGTGAAGAGAAAGAGTATTTGGCCCGAGAGAGCTTTCA 420
 Db 361 GACATCATGCTATGAAGTGAAGAGAAAGAGTATTTGGCCCGAGAGAGCTTTCA 420
 QY 421 TTTTGGAGAAAGAGCGAATATTTATCTCGAAGCAAGCCGTGATCCCAATTA 480
 Db 421 TTTTGGAGAAAGAGCGAATATTTATCTCGAAGCAAGCCGTGATCCCAATTA 480
 QY 481 CAGTATGCTTTTACGACAAAATACCTTTATCTGTCATGGAATATCACCTCGAGGG 540
 Db 481 CAGTATGCTTTTACGACAAAATACCTTTATCTGTCATGGAATATCACCTCGAGGG 540
 QY 541 GACTTGCTGCACTTTTGAATATATATGAGACACAGTATATGAAAACCTGATCACTTT 600
 Db 541 GACTTGCTGCACTTTTGAATATATATGAGACACAGTATATGAAAACCTGATCACTTT 600
 QY 601 TACCTAGCTGAGCTGATTTTGGCTGTCACAGCGTTCATCTGATGGATACGTGCATGA 660
 Db 601 TACCTAGCTGAGCTGATTTTGGCTGTCACAGCGTTCATCTGATGGATACGTGCATGA 660
 QY 661 GACATCAAGCTTGAAGACATTTCTGTTGACCGCACAGACATCAAGCTGTGATTTT 720
 Db 661 GACATCAAGCTTGAAGACATTTCTGTTGACCGCACAGACATCAAGCTGTGATTTT 720
 QY 721 GGAATCGCCGCGAAATATGATTCAAAACAAGATGTAATGCGAACTCCGATTTGGAGC 780
 Db 721 GGAATCGCCGCGAAATATGATTCAAAACAAGATGTAATGCGAACTCCGATTTGGAGC 780
 QY 781 CCAGATTAATAGCTCTCTGTAAGTGTGACTGTGATGAACGGAGATGAAAAGCACCTAC 840
 Db 781 CCAGATTAATAGCTCTCTGTAAGTGTGACTGTGATGAACGGAGATGAAAAGCACCTAC 840
 QY 841 GGCCTGACCTGTGACTGTGCTGATGAGGCGGTGATTTGCTATGATGATTTATGGAGA 900
 Db 841 GGCCTGACCTGTGACTGTGCTGATGAGGCGGTGATTTGCTATGATGATTTATGGAGA 900
 QY 901 TCCGCTTGCAGAGAGGAAACCTGTCAGAACCTTCAATTAACCTTTATTTCCAGCGG 960
 Db 901 TCCGCTTGCAGAGAGGAAACCTGTCAGAACCTTCAATTAACCTTTATTTCCAGCGG 960
 QY 961 TTTTGAATTTTCCAGATGACCCCAAGAGTGAAGTGAAGTCTTTGATGATGATTCAAAGC 1020
 Db 961 TTTTGAATTTTCCAGATGACCCCAAGAGTGAAGTGAAGTCTTTGATGATGATTCAAAGC 1020
 QY 1021 TTGTTGTGGGCGCAAAAGAGACTGAAGTGAAGTCTTTGCTGCAATCCTTTCTTC 1080
 Db 1021 TTGTTGTGGGCGCAAAAGAGACTGAAGTGAAGTCTTTGCTGCAATCCTTTCTTC 1080
 QY 1081 TCTAAATTAATGACTGGAACAACATCTGTAACCTCTCCGCTTCTGCTCCACCTCAG 1140
 Db 1081 TCTAAATTAATGACTGGAACAACATCTGTAACCTCTCCGCTTCTGCTCCACCTCAG 1140
 QY 1141 TCTGAGATGACACTTCCAAATTTTATGTAACCAAGAAATTCGTGAGTTTATCTCT 1200
 Db 1141 TCTGAGATGACACTTCCAAATTTTATGTAACCAAGAAATTCGTGAGTTTATCTCT 1200
 QY 1201 CCGTGTCACTGAGCCCTCAGAGCTTCTGGGTGAAGAACTGCGGTTTGTGGGTTTTCG 1260
 Db 1201 CCGTGTCACTGAGCCCTCAGAGCTTCTGGGTGAAGAACTGCGGTTTGTGGGTTTTCG 1260
 QY 1261 TACAGCAAGGCACTGGGAGATTCTTGATGATCTGAGTCTGTGTGTCGGGTCGAGCTCC 1320
 Db 1261 TACAGCAAGGCACTGGGAGATTCTTGATGATCTGAGTCTGTGTGTCGGGTCGAGCTCC 1320
 QY 1321 CCTGCCAAGACTAGCTTCATATGAAAAGAACTTCTCATCAAAAGCAAGAGCTACAAGAC 1380
 Db 1321 CCTGCCAAGACTAGCTTCATATGAAAAGAACTTCTCATCAAAAGCAAGAGCTACAAGAC 1380
 QY 1381 TCTCAGAGCAAGTGTCACAAGATGAGAGCAAGAAATGACCCGGTTACATGAGAGTGTCA 1440

Db 1381 TCTCAGGACAACTGTCAACAAGATGGAGCGAGAAATGACCCGGTTACATCGAGAGTGTCA 1440
QY 1441 GAGGTGGAGGCTGTGCTTGTGTCAAGAAAGGTTGAGCTGAAGGCTTGAAGCTCAAGAGA 1500
Db 1441 GAGGTGGAGGCTGTGCTTGTGTCAAGAAAGGTTGAGCTGAAGGCTTGAAGCTCAAGAGA 1500
QY 1501 TCCCTCTGGAGCAGAGACCTTGCTACATCAACAGAAATGCAATGCAATGCTTAAAGCAAGT 1560
Db 1501 TCCCTCTGGAGCAGAGACCTTGCTACATCAACAGAAATGCAATGCAATGCTTAAAGCAAGT 1560
QY 1561 TTGGAGCAAGACCGATGGAGGTGTCCAGAGAGATGCAAAAGCACTGACGCTTCCAT 1620
Db 1561 TTGGAGCAAGACCGATGGAGGTGTCCAGAGAGATGCAAAAGCACTGACGCTTCCAT 1620
QY 1621 GATATCAAGAGCAGAGACCGGAAAGCTCCAGAAATCAAGAGCAGAGTCCAGGCTCA 1680
Db 1621 GATATCAAGAGCAGAGACCGGAAAGCTCCAGAAATCAAGAGCAGAGTCCAGGCTCA 1680
QY 1681 GTGGAAGAAATGAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Db 1681 GTGGAAGAAATGAGGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 CGGAGTGATCTCTACGAATCTGAGCTGAGAGGTCTGGCTTGTCTGAGAAATTTCAAG 1800
Db 1741 CGGAGTGATCTCTACGAATCTGAGCTGAGAGGTCTGGCTTGTCTGAGAAATTTCAAG 1800
QY 1801 CGGAAAGCAGAGATGTCACATTAACCTGTTGAAGGCTTAAGGATCAAGGAGGCTGAA 1860
Db 1801 CGGAAAGCAGAGATGTCACATTAACCTGTTGAAGGCTTAAGGATCAAGGAGGCTGAA 1860
QY 1861 GTGGAGAAATATGAGAACTGAGAAAGATCAATGCTGAGCAGCTTCAAAATTCAGAG 1920
Db 1861 GTGGAGAAATATGAGAACTGAGAAAGATCAATGCTGAGCAGCTTCAAAATTCAGAG 1920
QY 1921 CTCCAGAGAACTGAGAAAGCTGTAAAGCCAGCACTGAGGCACTGAGCTGAG 1980
Db 1921 CTCCAGAGAACTGAGAAAGCTGTAAAGCCAGCACTGAGGCACTGAGCTGAG 1980
QY 1981 AATATCCGCGCAAGGAAAGGAGCCGAGAGGAGGAGCTGAGAAAGCTGCAAAACCGAG 2040
Db 1981 AATATCCGCGCAAGGAAAGGAGCCGAGAGGAGGAGCTGAGAAAGCTGCAAAACCGAG 2040
QY 2041 GATCTCTGAGAGCATCAAGAAAGAGTGTGAGCTGAGAGCCCGCATTTCTG 2100
Db 2041 GATCTCTGAGAGCATCAAGAAAGAGTGTGAGAGCTGAGAGCCCGCATTTCTG 2100
QY 2101 GAGAACAGGTTAAAGAGCTAGAGCCATGAGCGTAGAGAAACAGACTGAGAGTAC 2160
Db 2101 GAGAACAGGTTAAAGAGCTAGAGCCATGAGCGTAGAGAAACAGACTGAGAGTAC 2160
QY 2161 ATCCAGCAAAATCCCAACAGATCCAGCAAGTGGCTGATTAATTTGAGAGCTGAGAG 2220
Db 2161 ATCCAGCAAAATCCCAACAGATCCAGCAAGTGGCTGATTAATTTGAGAGCTGAGAG 2220
QY 2221 AAACATGGAGGCGCAAGTCTAGCCAGACCTAGAGCTGAAGTGAACCTGAAGCAAGAA 2280
Db 2221 AAACATGGAGGCGCAAGTCTAGCCAGACCTAGAGAGTGAACCTGAAGCAAGAA 2280
QY 2281 CAGCACTAGAGAAAGATTAAGTTGAGCAATCAGATTAAGAAAGCTGCTGAC 2340
Db 2281 CAGCACTAGAGAAAGATTAAGTTGAGCAATCAGATTAAGAAAGCTGCTGAC 2340
QY 2341 AAGGAGACACTGAGAAACATGATGAGAGCAGAGAGAGAGGCCATGAGAGGCAAA 2400
Db 2341 AAGGAGACACTGAGAAACATGATGAGAGCAGAGAGAGAGGCCATGAGAGGCAAA 2400
QY 2401 ATTCTCAGGCAACAGAGGAGATCAATGCTATGATTCAGAGATCAGATCCCTGGA 2460
Db 2401 ATTCTCAGGCAACAGAGGAGATCAATGCTATGATTCAGAGATCAGATCCCTGGA 2460
QY 2461 CAGAGATTGTGAACTGTCTGAAGCCAAATTAACCTGACAGAAATGACAGTCTTTTACC 2520
Db 2461 CAGAGATTGTGAACTGTCTGAAGCCAAATTAACCTGACAGAAATGACAGTCTTTTACC 2520

Db 2461 CAGAGATTGTGAACTGTCTGAAGCCAAATTAACCTGACAGAAATGACAGTCTTTTACC 2520
QY 2521 CAAAGAAACATGAAGGCCCAAGAGAGATGATTTCTGAATCAGGCAAGAAATTTTAC 2580
Db 2521 CAAAGAAACATGAAGGCCCAAGAGAGATGATTTCTGAATCAGGCAAGAAATTTTAC 2580
QY 2581 CTGAGACACAGGCTGGAAAGTTGAGAGGCCAGAAACCGAAACTGAGAGAGAGCTGAG 2640
Db 2581 CTGAGACACAGGCTGGAAAGTTGAGAGGCCAGAAACCGAAACTGAGAGAGAGCTGAG 2640
QY 2641 AAGATCAGCCACAAAGACCAAGATCAAGAAATCGGCTGCTGAACTGAGACCAAGATTG 2700
Db 2641 AAGATCAGCCACAAAGACCAAGATCAAGAAATCGGCTGCTGAACTGAGACCAAGATTG 2700
QY 2701 CGGAGGTCAAGCTGAGAGACAGAGAGAGAAACTGAGCTCAAGGCCAGCTCAAGAG 2760
Db 2701 CGGAGGTCAAGCTGAGAGACAGAGAGAGAAACTGAGCTCAAGGCCAGCTCAAGAG 2760
QY 2761 CTACAGCTCTCCCTGCAAGAGCGAGTCAAGTTGACAGCCCTGCAAGGCTCAAGGCT 2820
Db 2761 CTACAGCTCTCCCTGCAAGAGCGAGTCAAGTTGACAGCCCTGCAAGGCTCAAGGCT 2820
QY 2821 GCCCTGAGAGCCAGCTTCCGCAAGCCAGACAGAGCTGAGAGAGACCAAGAGAGCT 2880
Db 2821 GCCCTGAGAGCCAGCTTCCGCAAGCCAGACAGAGCTGAGAGAGACCAAGAGAGCT 2880
QY 2881 GAAAGAGATCCAGGCACTCAGGCAATGAGATGAAATCAGCGCAAAATTTGATGCT 2940
Db 2881 GAAAGAGATCCAGGCACTCAGGCAATGAGATGAAATCAGCGCAAAATTTGATGCT 2940
QY 2941 CTTGTAAACAGCTGTACTGTAAATCAACAGCTGAGAGAGAGCTTAAACAGCTGAG 3000
Db 2941 CTTGTAAACAGCTGTACTGTAAATCAACAGCTGAGAGAGAGCTTAAACAGCTGAG 3000
QY 3001 GACAAGCTGAACCTCAACCAAACTTTTACTTGTCCAAACAACTGATGAGGCTTCT 3060
Db 3001 GACAAGCTGAACCTCAACCAAACTTTTACTTGTCCAAACAACTGATGAGGCTTCT 3060
QY 3061 GCGCCCAACGAGATTTGTAACACTGAGAGGTAAGTGACCAATCTCCGCGGAGATC 3120
Db 3061 GCGCCCAACGAGATTTGTAACACTGAGAGGTAAGTGACCAATCTCCGCGGAGATC 3120
QY 3121 ACGGAACGAGAGATGAGCTTACCGACAGAGCAAAAGATGAGAGGCTTGAAGCAAG 3180
Db 3121 ACGGAACGAGAGATGAGCTTACCGACAGAGCAAAAGATGAGAGGCTTGAAGCAAG 3180
QY 3181 TGCAACATGCTGAGAGAAACAGCTATGATTTGAGAGGCTTAAACGATGAGCTGTA 3240
Db 3181 TGCAACATGCTGAGAGAAACAGCTATGATTTGAGAGGCTTAAACGATGAGCTGTA 3240
QY 3241 AAAGAGCGGAGTGGAGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCCAATTTGAG 3300
Db 3241 AAAGAGCGGAGTGGAGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCCAATTTGAG 3300
QY 3301 TGTGGGTTGAGAGCTGAGAGAGTGTGAGACCCAGAAACAGAGAGGCGAGAGCC 3360
Db 3301 TGTGGGTTGAGAGCTGAGAGAGTGTGAGACCCAGAAACAGAGAGGCGAGAGCC 3360
QY 3361 GATCAGCGATTCACCAAGTCTCGCCAGGTGTGAGAGCTGAGAGTGAAGAGCAAGGCT 3420
Db 3361 GATCAGCGATTCACCAAGTCTCGCCAGGTGTGAGAGCTGAGAGTGAAGAGCAAGGCT 3420
QY 3421 GAGATTCGCTCTGAGCAGGCTCTCAAGAGACAGAGCTGAAGGCGAGAGGCTCTCT 3480
Db 3421 GAGATTCGCTCTGAGCAGGCTCTCTCAAGAGACAGAGCTGAAGGCGAGAGGCTCTCT 3480
QY 3481 GACAACTCAATGACTGAGAAAGCAATGCTATCTTGAATGAAATGCCCCAGAGCTTA 3540
Db 3481 GACAACTCAATGACTGAGAAAGCAATGCTATCTTGAATGAAATGCCCCAGAGCTTA 3540
QY 3541 CAGCAGAGCTGAGAGCTGAACAGAGGCTCAACAGAGGCTTCTGAGAGAGCAAGCCAAA 3600
Db 3541 CAGCAGAGCTGAGAGCTGAACAGAGGCTTCTGAGAGAGCAAGCCAAA 3600

QY 3601 TTACAGCAGAGATGAGCTGCGAGAAAAATCAATTTCCGTGACTCAAGACTGCAA 3660
DB 3601 TTACAGCAGAGATGAGCTGCGAGAAAAATCAATTTCCGTGACTCAAGACTGCAA 3660
QY 3661 GAAGCTTAAATCGGGCTGATCTAAGTAAGACAGAAAAAGAGTGAATCTTGAGTACAGCTG 3720
DB 3661 GAAGCTTAAATCGGGCTGATCTAAGTAAGACAGAAAAAGAGTGAATCTTGAGTACAGCTG 3720
QY 3721 GAAACATTGAGTTCTCTATTTCTCATGAAAAGTGAAGTGAAGGCACTATTTCTCAA 3780
DB 3721 GAAACATTGAGTTCTCTATTTCTCATGAAAAGTGAAGTGAAGGCACTATTTCTCAA 3780
QY 3781 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCAACCTGTAAAAAGAAAAAG 3840
DB 3781 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCAACCTGTAAAAAGAAAAAG 3840
QY 3841 GTTCCCTGAGATGATGAGCTGAGAGTGGCCCTTGAGAGAGAAAGTCCCTGCTGCA 3900
DB 3841 GTTCCCTGAGATGATGAGCTGAGAGTGGCCCTTGAGAGAGAAAGTCCCTGCTGCA 3900
QY 3901 GAGCTGAGAGAAACCCCTTCAAGAAACCCGATCGAGTCCGCTCCGCGGAGAGAGCT 3960
DB 3901 GAGCTGAGAGAAACCCCTTCAAGAAACCCGATCGAGTCCGCTCCGCGGAGAGAGCT 3960
QY 3961 GCCCACCAGAAAGCAAGACCAACCCATCCACGCGACCAACCGGAGAGAGAG 4020
DB 3961 GCCCACCAGAAAGCAAGACCAACCCATCCACGCGACCAACCGGAGAGAGAG 4020
QY 4021 ATGCGCATGTCGCGCATCGTGGGATGCGGAGACACAGCCGATCGATGAGCTGCTG 4080
DB 4021 ATGCGCATGTCGCGCATCGTGGGATGCGGAGACACAGCCGATCGATGAGCTGCTG 4080
QY 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTCGCGCTTT 4140
DB 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTCGCGCTTT 4140
QY 4141 AAGGAAGCATGACACACAAATATTTCTCAACCGATTCAGAGTGAAGTGAAGAGCC 4200
DB 4141 AAGGAAGCATGACACACAAATATTTCTCAACCGATTCAGAGTGAAGTGAAGAGCC 4200
QY 4201 ACAAAAGTGTGCTGTGTGTGATACCGGATACCTTGAGACGGCAGGATCCAAATGTC 4260
DB 4201 ACAAAAGTGTGCTGTGTGTGATACCGGATACCTTGAGACGGCAGGATCCAAATGTC 4260
QY 4261 GAAATGTCAAGTATGTGTCAACCCCAAGTGTCTGACGTCCTTGACAGCCACTGCGGCTTG 4320
DB 4261 GAAATGTCAAGTATGTGTCAACCCCAAGTGTCTGACGTCCTTGACAGCCACTGCGGCTTG 4320
QY 4321 CCTGCTGAATATGCCCAACATTTCAACGAGGCTTTCTGCGGTGACAAATGAACTCCCA 4380
DB 4321 CCTGCTGAATATGCCCAACATTTCAACGAGGCTTTCTGCGGTGACAAATGAACTCCCA 4380
QY 4381 GGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACTTGAAAGGATGATGAAGTGCC 4440
DB 4381 GGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACTTGAAAGGATGATGAAGTGCC 4440
QY 4441 AAGAAATTAACAAACAGAGACAGAGGCTGGACAGAGAAATGATGCTCTGAGAGGATCA 4500
DB 4441 AAGAAATTAACAAACAGAGACAGAGGCTGGACAGAGAAATGATGCTCTGAGAGGATCA 4500
QY 4501 AAAAGTCCATTTAATGACAAATGAGCAGAGAGCTGAGACAGAGGCTGGAGAGATTT 4560
DB 4501 AAAAGTCCATTTAATGACAAATGAGCAGAGAGCTGAGACAGAGGCTGGAGAGATTT 4560
QY 4561 GAGCTGTGCTTCCGAGCGGGATGATATTTCAATGAGTGGCTGGTGGTCTTCCGAATCC 4620
DB 4561 GAGCTGTGCTTCCGAGCGGGATGATATTTCAATGAGTGGCTGGTGGTCTTCCGAATCC 4620
QY 4621 GCAATTAACAGCCAAAGAGATGTCCCATATACATGAGATGAAATTCACCCGACACCC 4680
DB 4621 GCAATTAACAGCCAAAGAGATGTCCCATATACATGAGATGAAATTCACCCGACACCC 4680

QY 4681 ACCGTGTGGCCCGGAGAAACCTCTATCTTGCTAGCTCCAGCTTCCCTGACAAACAGCGC 4740
DB 4681 ACCGTGTGGCCCGGAGAAACCTCTATCTTGCTAGCTCCAGCTTCCCTGACAAACAGCGC 4740
QY 4741 TGGGTACCGCCTTAAGATCAATTTGTCAGAGTGGAGAGATTTCTAGAGAAAAAGCAAA 4800
DB 4741 TGGGTACCGCCTTAAGATCAATTTGTCAGAGTGGAGAGATTTCTAGAGAAAAAGCAAA 4800
QY 4801 GCTGATGCTAAATGCTTGGAATCTCCCTGCTGAACTGGAAGTGTATGACCTGTAGAC 4860
DB 4801 GCTGATGCTAAATGCTTGGAATCTCCCTGCTGAACTGGAAGTGTATGACCTGTAGAC 4860
QY 4861 ATGAATGCAAGCTGCTTCACTGACCAAGTGTGTGTTGGGACCCGAGAGAGGCTC 4920
DB 4861 ATGAATGCAAGCTGCTTCACTGACCAAGTGTGTGTTGGGACCCGAGAGAGGCTC 4920
QY 4921 TACGCCCTGAATGTCTTGAATAAATCTCCCTAACCATGTCCAGAAATTTGAGCAGTTC 4980
DB 4921 TACGCCCTGAATGTCTTGAATAAATCTCCCTAACCATGTCCAGAAATTTGAGCAGTTC 4980
QY 4981 CAAATTTATATATCAAGGACCTGAGAGCTACTATGATAGCAGAGAGAGCGGCA 5040
DB 4981 CAAATTTATATATCAAGGACCTGAGAGCTACTATGATAGCAGAGAGAGCGGCA 5040
QY 5041 CTGTGTCTTGTGACGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCC 5100
DB 5041 CTGTGTCTTGTGACGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCC 5100
QY 5101 CAGCCGACATCTCACCCCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGGGGCA 5160
DB 5101 CAGCCGACATCTCACCCCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGGGGCA 5160
QY 5161 GGCAGATTGAGACGGGCTGTGATCTGTGACAGCATGCCGAGAAAGTCGATCTC 5220
DB 5161 GGCAGATTGAGACGGGCTGTGATCTGTGACAGCATGCCGAGAAAGTCGATCTC 5220
QY 5221 CGCTACAGAAACCTCGACAAATCTGATCCGAGAAAGATAGAGACTTCAGAGCC 5280
DB 5221 CGCTACAGAAACCTCGACAAATCTGATCCGAGAAAGATAGAGACTTCAGAGCC 5280
QY 5281 TCGAGCTGATATCACTTCCAAATTAAGATTCCTATTTGGAACCAATTAATTTAGAA 5340
DB 5281 TCGAGCTGATATCACTTCCAAATTAAGATTCCTATTTGGAACCAATTAATTTAGAA 5340
QY 5341 ATGACATAGACAGACACAGCTCGAGAGATTCCTGATTAAGATAGCAATTCCTGGCA 5400
DB 5341 ATGACATAGACAGACACAGCTCGAGAGATTCCTGATTAAGATAGCAATTCCTGGCA 5400
QY 5401 CCTGTGTGTGTCGCGCTCTTCCAAACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAG 5460
DB 5401 CCTGTGTGTGTCGCGCTCTTCCAAACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAG 5460
QY 5461 GAGAGGACGAGAGAGAGTACTGCTGTGTTCACAGAAATTTGAGATGTTCTGAGATTT 5520
DB 5461 GAGAGGACGAGAGAGAGTACTGCTGTGTTCACAGAAATTTGAGATGTTCTGAGATTT 5520
QY 5521 TACGAGAAAGCTTAGCCGACAGACAGATCTCAAGTGAAGTGGCTTACCTTTGCTG 5580
DB 5521 TACGAGAAAGCTTAGCCGACAGACAGATCTCAAGTGAAGTGGCTTACCTTTGCTG 5580
QY 5581 TACAGAGAAACCTATCTGTTTGTGAACCACTTCAACTCACTGGAAGTATTAAGATCCAG 5640
DB 5581 TACAGAGAAACCTATCTGTTTGTGAACCACTTCAACTCACTGGAAGTATTAAGATCCAG 5640
QY 5641 GAGAGCTCTCGAGAGAGACCCCTGCGGAGGTACTGAGACATCCCGAACCCGCGCTAC 5700
DB 5641 GAGAGCTCTCGAGAGAGACCCCTGCGGAGGTACTGAGACATCCCGAACCCGCGCTAC 5700
QY 5701 CTGGGCTCTGACATTTCTCAGAGAGATTTACTTGGGCTCTCATTAACAGAGATTAATTA 5760
DB 5701 CTGGGCTCTGACATTTCTCAGAGAGATTTACTTGGGCTCTCATTAACAGAGATTAATTA 5760
QY 5761 AAGGTCAATTTGCTGCAAGGAAACCTGCTGAAGAGTCCGAGCACTGAACACCAACCGG 5820

Db 5761 AGGTCATTCTGCTGAAGGAAACCTGTAAGAGTCCGGACATGAACACACCGGGGC 5820
 Qy 5821 CCGTCCACCTCCCGGAGAGGCCCCAAGAGAGGAGCCCAAGCTTACAGAGACATC 5880
 Db 5821 CCGTCCACCTCCCGGAGAGGCCCCAAGAGAGGAGCCCAAGCTTACAGAGACATC 5880
 Qy 5881 ACCAAGGCGGCTGCTCCAGGCCAGCGCCCGGAGAGGCCCCAAGCTTACAGAGACATC 5940
 Db 5881 ACCAAGGCGGCTGCTCCAGGCCAGCGCCCGGAGAGGCCCCAAGCTTACAGAGACATC 5940
 Qy 5941 AGCAGACCCCAACCGCTACCGGAGAGGCGGAGCCAGCTGCGAGAGCAAACTTCTCTGAC 6000
 Db 5941 AGCAGACCCCAACCGCTACCGGAGAGGCGGAGCCAGCTGCGAGAGCAAACTTCTCTGAC 6000
 Qy 6001 CGCCCTTGGAGAGAGAGAGTCCCCCGGCTGCTTACAGACAGCGAGAGAGCGGTCC 6060
 Db 6001 CGCCCTTGGAGAGAGAGAGTCCCCCGGCTGCTTACAGACAGCGAGAGAGCGGTCC 6060
 Qy 6061 CCGGAGAGCTGTTTGAAGACAGAGAGAGGCGGCTGCTGCGAGAGCGGTGAGACC 6120
 Db 6061 CCGGAGAGCTGTTTGAAGACAGAGAGAGGCGGCTGCTGCGAGAGCGGTGAGACC 6120
 Qy 6121 CCGCTGTCCAGGTGAACAAAGTCTGGACCAAGTCTTCAATATA 6165
 Db 6121 CCGCTGTCCAGGTGAACAAAGTCTGGACCAAGTCTTCAATATA 6165

RESULT 5
 AAL55217
 ID AAL55217 standard; DNA; 6156 BP.
 AC AAL55217;
 XX

DT 01-MAY-2003 (first entry)
 XX
 DE Human CR1K related DNA sequence, SEQ ID No 8.
 XX

XX Anorectic; hypotensive; cardiatic; antilipemic; cerebroprotective;
 KW antiobesity; osteoporotic; antidiabetic; cytoskeletal; antidepressant;
 KW immunomodulator; antidiabetic; antiparkinsonian; neurotrophic;
 KW neuroprotective; antiinflammatory; antidiabetic; analgesic;
 KW human citron rho/rac-interacting kinase; enzyme; CR1K; ameliorating;
 KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;
 KW central nervous system disorder; chronic obstructive pulmonary disease;
 KW diabetes; pain; ds.

XX Homo sapiens.
 OS

XX WO2003004523-A1.
 PN

XX 16-JAN-2003.
 PD

XX 28-JUN-2002; 2002WO-BP07156.
 PF

XX 02-JUL-2001; 2001US-301841P.
 PR

XX 11-DEC-2001; 2001US-338611P.
 PR

XX 25-APR-2002; 2002US-375014P.
 PR

XX (FARB) BAYER AG.
 PA

XX Zhu Z;
 PI

XX WPI; 2003-221576/21.
 DR

XX New human citron rho/rac-interacting kinase (CR1K) polypeptide and
 PT polynucleotide, useful in preventing, ameliorating or treating diseases
 PT Alzheimer's disease -

XX Disclosure; Page 217-222; 237pp; English.
 PS

XX The invention relates to an isolated polynucleotide encoding a human
 CC

CC citron rho/rac-interacting kinase polypeptide. The isolated
 CC polynucleotide comprises a 6156 or 8603 base pair sequence, given in the
 CC specification. The human citron rho/rac-interacting kinase (CR1K)
 CC polypeptide and polynucleotide are useful in preventing, ameliorating, or
 CC treating diseases associated with human CR1K dysfunction such as obesity
 CC and obesity-associated comorbidities (e.g. hypertension, coronary artery
 CC disease, hyperlipidemia, stroke, gout, osteoarthritis, some types of
 CC cancer including endometrial, breast, prostate and colon cancer),
 CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood
 CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's
 CC disease), chronic obstructive pulmonary disease, or diabetes. These can
 CC also be used to treat pain associated with the disorders. The human CR1K
 CC polypeptide is also useful in diagnostic assays or in genetic testing.
 CC The expression vector or the reagent is useful in preparing a medicament
 CC for modulating the activity of a human CR1K in a disease, e.g. obesity, a
 CC central nervous system disorder, or chronic obstructive pulmonary
 CC disease. The fusion protein is useful for generating antibodies against a
 CC CR1K polypeptide and for use in various assay systems. The methods are
 CC useful in producing and detecting the polynucleotide and polypeptide and
 CC in screening for agents that modulate the activity of the human CR1K
 CC polypeptide. This polynucleotide sequence represents a DNA sequence
 CC relating to the human CR1K protein of the invention.
 CC

SQ Sequence 6156 BP; 1732 A; 1548 C; 1679 G; 1197 T; 0 other;
 Query Match 99.6%; Score 6142.2; DB 25; Length 6156;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 6147; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGTTGAAGTTCAAAATATGAGAGCGGAATCTTTGGATGCTGCTGCTGAACCAT 60
 Db 1 ATGTTGAAGTTCAAAATATGAGAGCGGAATCTTTGGATGCTGCTGCTGAACCAT 60
 Qy 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCCAGAGGGAACCAACCTTTATGACTCAA 120
 Db 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCCAGAGGGAACCAACCTTTATGACTCAA 120
 Qy 121 CAGCAGATGCTCTCTTCTCCGAGAGAGATATAGATCCCTCTTGTCTTGTGA 180
 Db 121 CAGCAGATGCTCTCTTCTCCGAGAGAGATATAGATCCCTCTTGTCTTGTGA 180
 Qy 181 GAATGAGTCAAGCTCTCTGATGAAGATTAAGACAGTGAACATTTGTCCGGAATAT 240
 Db 181 GAATGAGTCAAGCTCTCTGATGAAGATTAAGACAGTGAACATTTGTCCGGAATAT 240
 Qy 241 TCCGACACCATAGCTGATGTTAAGAGGCTCCAGGCTTCCGAGAAAGATTCAGATCAG 300
 Db 241 TCCGACACCATAGCTGATGTTAAGAGGCTCCAGGCTTCCGAGAAAGATTCAGATCAG 300
 Qy 301 AGCTTTGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 Db 301 AGCTTTGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 Qy 361 GACATCTATGCTATGAAGATGAAGAAAGGCTTTATTTGGCCAGAGAGAGGTTTCA 420
 Db 361 GACATCTATGCTATGAAGATGAAGAAAGGCTTTATTTGGCCAGAGAGAGGTTTCA 420
 Qy 421 TTTTGGAGAGAGGAGGAGCAATATATTCAGAGAGCAAGCCCGGATCCCCCAATTA 480
 Db 421 TTTTGGAGAGAGGAGGAGCAATATATTCAGAGAGCAAGCCCGGATCCCCCAATTA 480
 Qy 481 CAGTATGCTTTCAAGACAAAATCACTTTATCTGATGATGAATATAGCTGAGAGG 540
 Db 481 CAGTATGCTTTCAAGACAAAATCACTTTATCTGATGATGAATATAGCTGAGAGG 540
 Qy 541 GACTTGCTGCTCTTTTGAATATGATGAGAGCCGTTATGATGAAGAAACCTGATACAGTT 600
 Db 541 GACTTGCTGCTCTTTTGAATATGATGAGAGCCGTTATGATGAAGAAACCTGATACAGTT 600
 Qy 601 TACCTAGCTGAGCTATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Db 601 TACCTAGCTGAGCTATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660


```

QY 5041 CTGTGTTCTTGAGACGGTAAGAAAGTAAACAGTCCCTGGCCCAAGTCCCACTTGCTGCC 5100
DB 5041 CTGTGTTCTTGAGACGGTAAGAAAGTAAACAGTCCCTGGCCCAAGTCCCACTTGCTGCC 5100
QY 5101 CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTTGGGGCA 5160
DB 5101 CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTTGGGGCA 5160
QY 5161 GGCAGATTGAGAACGGGCTGTGATCTGTGACGACATGCCAGCAAGTGGTCAATCTC 5220
DB 5161 GGCAGATTGAGAACGGGCTGTGATCTGTGACGACATGCCAGCAAGTGGTCAATCTC 5220
QY 5221 CGGTACAGAAAACCTCGCAAAATCTGATCCGGAAGAGATAGAGACTTAGAGCCC 5280
DB 5221 CGGTACAGAAAACCTCGCAAAATCTGATCCGGAAGAGATAGAGACTTAGAGCCC 5280
QY 5281 TCGAGCTGATCACTTCAACCAATTCAGATCTCTATTGGAACCAATTAATCTACGA 5340
DB 5281 TCGAGCTGATCACTTCAACCAATTCAGATCTCTATTGGAACCAATTAATCTACGA 5340
QY 5341 ATCGACATGAGACGTAACGCTCGAGAAATTCCTGATTAAGATGACCATTCCTTGGA 5400
DB 5341 ATCGACATGAGACGTAACGCTCGAGAAATTCCTGATTAAGATGACCATTCCTTGGA 5400
QY 5401 CCTGCTGTGTTGGCCCTCTTCCAAACAGTCCCTGTCATCTGTAAGGTGAACAGC 5460
DB 5401 CCTGCTGTGTTGGCCCTCTTCCAAACAGTCCCTGTCATCTGTAAGGTGAACAGC 5460
QY 5461 GCAGGGCAGCGAAGAGTACTTGTCTTCCAGAGCTTCCCTGTCATCTGTAAGGTGAACAGC 5520
DB 5461 GCAGGGCAGCGAAGAGTACTTGTCTTCCAGAGCTTCCCTGTCATCTGTAAGGTGAACAGC 5520
QY 5521 TACGGAAGCTGTAAGCCGACAGACATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTTGCC 5580
DB 5521 TACGGAAGCTGTAAGCCGACAGACATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTTGCC 5580
QY 5581 TACAGAGAACCTATCTGTTGTGAACCCACTTCAATCTGTAAGTATTTGATCCAG 5640
DB 5581 TACAGAGAACCTATCTGTTGTGAACCCACTTCAATCTGTAAGTATTTGATCCAG 5640
QY 5641 GCAGCCTCTCTAGCAGGAGCCCTGCGGAGCTACTTGAATCCGGAACCCGCGCTAC 5700
DB 5641 GCAGCCTCTCTAGCAGGAGCCCTGCGGAGCTACTTGAATCCGGAACCCGCGCTAC 5700
QY 5701 CTGGGCTCTGCACTTCTCTCAAGAGATTTACTTGGCTCTCTCAATCCAGATTAATTA 5760
DB 5701 CTGGGCTCTGCACTTCTCTCAAGAGATTTACTTGGCTCTCTCAATCCAGATTAATTA 5760
QY 5761 AGGTCATTGCTGCAAGGAAACCTCGTGAAGAGTCCGCACTGAAACCAACCGGGGC 5820
DB 5761 AGGTCATTGCTGCAAGGAAACCTCGTGAAGAGTCCGCACTGAAACCAACCGGGGC 5820
QY 5821 CCGTCCACCTCCCGCAGAGCCCAACGAGGCGCAACCGATACACGAGCAATC 5880
DB 5821 CCGTCCACCTCCCGCAGAGCCCAACGAGGCGCAACCGATACACGAGCAATC 5880
QY 5881 ACCAAGGGGTGGGCTTCCAGCCCAAGCGCGCCGCAAGGCGCCAGCCACCCGAGAGCCA 5940
DB 5881 ACCAAGGGGTGGGCTTCCAGCCCAAGCGCGCCGCAAGGCGCCAGCCACCCGAGAGCCA 5940
QY 5941 AGCAGACCCCAACCGCTACCGGAGGGGAGCCGAGCTGCGCAGGAGCAAGTCTCTGGC 6000
DB 5941 AGCAGACCCCAACCGCTACCGGAGGGGAGCCGAGCTGCGCAGGAGCAAGTCTCTGGC 6000
QY 6001 CGCCCTCTGAGAGCGAGAGAGTCCCGCGCGGATACTCAGACCGCGAGAGAGCGTCC 6060
DB 6001 CGCCCTCTGAGAGCGAGAGAGTCCCGCGCGGATACTCAGACCGCGAGAGAGCGTCC 6060
QY 6061 CCCGAGAGCTGTTTGAAGACAGCAGGAGGCGCGCTGCTGCGGAGCCGCTGAGACC 6120
DB 6061 CCCGAGAGCTGTTTGAAGACAGCAGGAGGCGCGCTGCTGCGGAGCCGCTGAGACC 6120
QY 6121 CCGCTGTCTCCAGGTGAACAGGTCTGGAGCCAGTC 6155

```

```

DB 6121 CCGCTGTCTCCAGGTGAACAGGTCTGGAGCCAGTC 6155
RESULT 6
ID AAS06701 standard; cDNA; 6159 BP.
XX AAS06701;
XX
DE 12-SEP-2001 (first entry)
XX
DE Polynucleotide sequence encoding human protein kinase #1.
XX
DE Human; protein kinase; PK; STK; cancer; cardiovascular disease;
XX metabolic disorder; immune related disease; neurological disorder;
XX neurodegenerative disorder; inflammatory disorder; infectious disease;
XX reproductive disorder; gene therapy; ss.
XX Homo sapiens.
XX
XX MO200138503-A2.
XX
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000MO-US32085.
XX
XX 24-NOV-1999; 99US-0167482.
XX
XX (SUBE-) SUGEN INC.
XX
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX Flanagan P, Clary D;
XX
XX WPI; 2001-343950/36.
XX P-PSDB; AAU03501.
XX
XX Nucleic acids encoding human kinase polypeptides, useful for preventing
XX diagnosing and/or treating e.g. cancer, immune, cardiovascular and
XX neuronal-associated diseases, and microbial infections -
XX
XX Example 1; Figure 1; 433bp; English.
XX
XX AAS06701-AAS06757 encode for novel human protein kinases #1-57. The
XX novel protein kinases have been identified as members of the tyrosine
XX or serine/threonine kinase (PTK and STK) families. The polynucleotides
XX encoding protein kinases and the polypeptides may be used in the
XX prevention, diagnosis and treatment of diseases associated with
XX cancers (especially cancers of hematopoietic origin), cardiovascular
XX disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
XX immune related diseases (e.g. rheumatoid arthritis), neurological
XX disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
XX Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
XX disease (e.g. HIV) and reproductive disorders (e.g. infertility).
XX Additionally, polynucleotides encoding protein kinases may be
XX used for gene therapy and as DNA probes in diagnostic assays.
XX The protein kinase polypeptides may be used as antigens in the production
XX of antibodies against the protein kinases and in assays to identify
XX modulators of protein kinase expression and activity.
XX
XX Sequence 6159 BP; 1732 A; 1549 C; 1680 G; 1198 T; 0 other;
SQ
Query Match 99.4%; Score 6126; DB 22; Length 6159;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 6145; Conservative 0; Mismatches 10; Indels 3; Gaps 1;
QY 1 ATGTTGAAGTTCAATATGAGACCGGATCTTGTGATGCTGTGATGACCCATT 60
DB 1 ATGTTGAAGTTCAATATGAGACCGGATCTTGTGATGCTGTGATGACCCATT 60
QY 61 GCCAGCCGGGCTCCAGGCTTAATCTGTTCTTCCAGGGGAAACCACTTTATGATCA 120

```

Dh 61 GCCAGCGGCGCTCCAGGCTGMAATCTGTCTCCAGGGGAAACCACTTTATAGTCMA 120
Qy 121 CAGAGATGTCCTCTCTTCCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180
Dh 121 CAGAGATGTCCTCTCTTCCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180
Qy 181 GAATGACATGACCTGCTGTAGTAAATTAAAGCAGTGAACAATTGTCCGAAA--G 237
Dh 181 GAATGACATGACCTGCTGTAGTAAATTAAAGCAGTGAACAATTGTCCGAAAAGT 240
Qy 238 TATTCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAAAGACCTTGAAGTC 297
Dh 241 TATTCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAAAGACCTTGAAGTC 300
Qy 298 AGAGCTCTGTAGGTGTGTCATCTTGTGTAAGTGAAGTGGTAAAGAGAACACC 357
Dh 301 AGAGCTCTGTAGGTGTGTCATCTTGTGTAAGTGGTAAAGAGAACACC 360
Qy 358 GGGGACATCTATGCTTATGAAGTGAAGAGAGGCTTTATTTGGCCGAGAGCAGTT 417
Dh 361 GGGGACATCTATGCTTATGAAGTGAAGAGAGGCTTTATTTGGCCGAGAGCAGTT 420
Qy 418 TCATTTTGAAGAGAGCGGAAACATATATCTCGAAGCACAAGCCCGTGGATCCGCCAA 477
Dh 421 TCATTTTGAAGAGAGCGGAAACATATATCTCGAAGCACAAGCCCGTGGATCCGCCAA 480
Qy 478 TTACAGTATGCTTTCAAGACAAATAATCACTTTATCTGTCATGGAATACAGCTTGA 537
Dh 481 TTACAGTATGCTTTCAAGACAAATAATCACTTTATCTGTCATGGAATATCAGCTTGA 540
Qy 538 GGGGACCTGCTGCTCACTTTGAATAGATAGAGGACCAAGTTAGTAAACCTGATACAG 597
Dh 541 GGGGACCTGCTGCTCACTTTGAATAGATAGAGGACCAAGTTAGTAAACCTGATACAG 600
Qy 598 TTTTACCTAGCTGATTTTGGCTGTTCAAGCCGTTCATCTGATGGATACGTGAT 657
Dh 601 TTTTACCTAGCTGATTTTGGCTGTTCAAGCCGTTCATCTGATGGATACGTGAT 660
Qy 658 CGAGACATCAAGCTTGAGAACTTCTGTTGACCGCAGAGACATCAAGCTGTGTGAT 717
Dh 661 CGAGACATCAAGCTTGAGAACTTCTGTTGACCGCAGAGACATCAAGCTGTGTGAT 720
Qy 718 TTTGATCTGCGCGAATAATGAATTCAAACAGATGTGTAATGCCAACTCCGATTGG 777
Dh 721 TTTGATCTGCGCGAATAATGAATTCAAACAGATGTGTAATGCCAACTCCGATTGG 780
Qy 778 ACCCAGATTACATGCTCTCTGAAAGTGTGTAATGTAAGAGAGAGAGAGAGAGAG 837
Dh 781 ACCCAGATTACATGCTCTCTGAAAGTGTGTAATGTAAGAGAGAGAGAGAGAGAG 840
Qy 838 TAGGCGCTGACCTGTGATGCTGTGGTCACTGTGGGTGATTTGCCATATAGATGATTTAGG 897
Dh 841 TAGGCGCTGACCTGTGATGCTGTGGTCACTGTGGGTGATTTGCCATATAGATGATTTAGG 900
Qy 898 AGATCCCTCTGCGAGAGGAACTCTGCGAGAACCTTCAATAACATTAATGAATTTCCAG 957
Dh 901 AGATCCCTCTGCGAGAGGAACTCTGCGAGAACCTTCAATAACATTAATGAATTTCCAG 960
Qy 958 CGGTTTTTGAATTTCCAGATGACCCCAAGTGAAGAGTCTTTCTGATCTGATTTCAA 1017
Dh 961 CGGTTTTTGAATTTCCAGATGACCCCAAGTGAAGAGTCTTTCTGATCTGATTTCAA 1020
Qy 1018 AGCTTGTGTGCGGCGAGAGAGAGAGAGAGAGTGAAGTCTTTGCTGCCATCTTTC 1077
Dh 1021 AGCTTGTGTGCGGCGAGAGAGAGAGAGAGTGAAGTCTTTGCTGCCATCTTTC 1080
Qy 1078 TTTCTTAAATTTGATGGAACAAATTCGTAATCTCTCTCCCTCTTCTTCCACCTTC 1137
Dh 1081 TTTCTTAAATTTGATGGAACAAATTCGTAATCTCTCTCCCTCTTCTTCCACCTTC 1140
Qy 1138 AAGCTGACATGACCTCCAAATTTTGAAGACAGAGAGAGAGATTCGTGGTTTATCC 1197
Dh 1141 AAGCTGACATGACCTCCAAATTTTGAAGACAGAGAGAGATTCGTGGTTTATCC 1200

Qy 1198 TCTCCGTGACAGCTAGGCCCCCTCAGGCTTCTCGGGTGAAGAATCGGTTTGTGGGGTTT 1257
Dh 1201 TCTCCGTGACAGCTAGGCCCCCTCAGGCTTCTCGGGTGAAGAATCGGTTTGTGGGGTTT 1260
Qy 1258 TCGTACAGCAAGCACTGAGGATTTCTTGTAGATCTGAGTCTGTGTGTGCGGTCTGAGC 1317
Dh 1261 TCGTACAGCAAGCACTGAGGATTTCTTGTAGATCTGAGTCTGTGTGTGCGGTCTGAGC 1320
Qy 1318 TCCCTGCGCAACATAGCTCCATGSAAGAAACCTTCTATCAAAAGCAAGAGAGCTTACA 1377
Dh 1321 TCCCTGCGCAACATAGCTCCATGSAAGAAACCTTCTATCAAAAGCAAGAGAGCTTACA 1380
Qy 1378 GACTCTCAGAGCAATGTCAACAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Dh 1381 GACTCTCAGAGCAATGTGTCAACAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1438 TCAAGAGTGAAGAGCTGTGCTTATGTCAGAAAGAGTGAAGAGTGAAGAGCTTGAAGCTCAG 1497
Dh 1441 TCAAGAGTGAAGAGCTGTGCTTATGTCAGAAAGAGTGAAGAGTGAAGAGCTTGAAGCTCAG 1500
Qy 1498 AGATCCCTCTGAG 1557
Dh 1501 AGATCCCTCTGAG 1560
Qy 1558 AGTTTGAAGCAAGCAAG 1617
Dh 1561 AGTTTGAAGCAAGCAAG 1620
Qy 1618 CATGATTCAG 1677
Dh 1621 CATGATTCAG 1680
Qy 1678 CAAGTGAAG 1737
Dh 1681 CAAGTGAAG 1740
Qy 1738 AGAGAGATGATCTTACAGATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGAGATTC 1797
Dh 1741 AGAGAGATGATCTTACAGATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGAGATTC 1800
Qy 1798 AAGCGGAAAG 1857
Dh 1801 AAGCGGAAAG 1860
Qy 1858 GAAAGTGAAG 1917
Dh 1861 GAAAGTGAAG 1920
Qy 1918 GAGCTCCAAAG 1977
Dh 1921 GAGCTCCAAAG 1980
Qy 1978 CAGAAATATCCGCAAG 2037
Dh 1981 CAGAAATATCCGCAAG 2040
Qy 2038 GAGGATTTCTTCTGAAGAGATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2097
Dh 2041 GAGGATTTCTTCTGAAGAGATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2098 CTGAG 2157
Dh 2101 CTGAG 2160
Qy 2158 GACATTCAGACAAATTCACAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2217
Dh 2161 GACATTCAGACAAATTCACAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2218 GAGAAACATCGGAG 2277
Dh 2221 GAGAAACATCGGAG 2280

QY 2278 GAGCAGACTATGAGGAAAAAGATTAAAGTGTGACATCATGATTAAGAAAGACTGGCT 2337
DB 2281 GACACAGACTATAGGAAAAAGATTAAAGTGTGACATCATGATTAAGAAAGACTGGCT 2340
QY 2338 GACAAAGAGACATGAGAAACATGATGCAGACACGAGAGAGAGGCCCATGAGAAAGGCG 2397
DB 2341 GAAAGAGAGACATGAGAAACATGATGCAGACACGAGAGAGAGGCCCATGAGAAAGGCG 2400
QY 2398 AAAATTCTGAGCAACAGAGGCGATGATCAATGCTATGGAATTCAGAGATCCAGTCC 2457
DB 2401 AAAATTCTGAGCAACAGAGGCGATGATCAATGCTATGGAATTCAGAGATCCAGTCC 2460
QY 2458 GAACAGAGATTGTGAACTGTCTGAAAGCAATTAACCTTGACCAATATGCGACTTTT 2517
DB 2461 GAACAGAGATTGTGAACTGTCTGAAAGCAATTAACCTTGACCAATATGCGACTTTT 2520
QY 2518 ACCCAAGAGACATGAGAGGCCCAAGAAAGATGATTTCTGAACCTGAGGACAGAAATTT 2577
DB 2521 ACCCAAGAGACATGAGAGGCCCAAGAAAGATGATTTCTGAACCTGAGGACAGAAATTT 2580
QY 2578 TACCTGAGACACAGGCTGGAAGTTGGAGGCCAGAACCCGAAACTGGAGAGCAGCTG 2637
DB 2581 TACCTGAGACACAGGCTGGAAGTTGGAGGCCAGAACCCGAAACTGGAGAGCAGCTG 2640
QY 2638 GAGAAGATCAGCCACCAAGACCAAGTGAACAAGATCCGCTGCTGGAACCTGAGACAA 2697
DB 2641 GAGAAGATCAGCCACCAAGACCAAGTGAACAAGATCCGCTGCTGGAACCTGAGACAA 2700
QY 2698 TTGCGGAGAGTCAGTCTAGAGACAGAGAGCAAGAACTGGAGCTCAAGGCCAGCTCACA 2757
DB 2701 TTGCGGAGAGTCAGTCTAGAGACAGAGAGCAAGAACTGGAGCTCAAGGCCAGCTCACA 2760
QY 2758 GAGCTACAGCTCTCCCTGCAAGAGCGCGAGTCAAGTTGACAGCCCTGAGGCTGCA 2817
DB 2761 GAGCTACAGCTCTCCCTGCAAGAGCGCGAGTCAAGTTGACAGCCCTGAGGCTGCA 2820
QY 2818 GCGGCTCTGAGAGAGCAGCTTGCGCAGAGGAGAGAGAGCTGGAGAGAGCCACACAGAA 2877
DB 2821 GCGGCTCTGAGAGAGCAGCTTGCGCAGAGGAGAGAGAGCTGGAGAGAGCCACACAGAA 2880
QY 2878 GCTGAAGAGAGATCCAGGCACTACGACATAGAGATGAATTCAGGCAATTTTGAT 2937
DB 2881 GCTGAAGAGAGATCCAGGCACTACGACATAGAGATGAATTCAGGCAATTTTGAT 2940
QY 2938 GCTCTTCTGAACAGCTGTACTGTATATCAAGCTGAGAGAGAGCTTAACCAAGCTGAC 2997
DB 2941 GCTCTTCTGAACAGCTGTACTGTATATCAAGCTGAGAGAGAGCTTAACCAAGCTGAC 3000
QY 2998 GAGGACAAAGCTGAACCTCAACCAACAACTTCTACTGTCCAACCAACTGATGAGGCT 3057
DB 3001 GAGGACAAAGCTGAACCTCAACCAACAACTTCTACTGTCCAACCAACTGATGAGGCT 3060
QY 3058 TCTGCGCGCAACGACGAGATTGTACACTGCGAAGTGAAGTGAACATCTCCGCGGAG 3117
DB 3061 TCTGCGCGCAACGACGAGATTGTACACTGCGAAGTGAAGTGAACATCTCCGCGGAG 3120
QY 3118 ATCAAGAGACAGAGATGACAGCTTACAGCCAGAAAGAAACGATGAGAGGCTCTGAAGCC 3177
DB 3121 ATCAAGAGACAGAGATGACAGCTTACAGCCAGAAAGAAACGATGAGAGGCTCTGAAGCC 3180
QY 3178 ACGTGACCAATGCTGAGAGAAACAGGTCAATGATTGAGGCTTAACGATGAGGTGTA 3237
DB 3181 ACGTGACCAATGCTGAGAGAAACAGGTCAATGATTGAGGCTTAACGATGAGGTGTA 3240
QY 3238 GAAAAAGAGCGGCACTGAGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCAGTTT 3297
DB 3241 GAAAAAGAGCGGCACTGAGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCAGTTT 3300
QY 3298 GAGGTGCGGTTTGAAGAGCTGACAGAGATGCTGAGACCCGAGAAACAGAGAGGCGAG 3357
DB 3301 GAGGTGCGGTTTGAAGAGCTGACAGAGATGCTGAGACCCGAGAAACAGAGAGGCGAG 3360
QY 3358 GCCGATCAGCGGATCACCGAGTCTGCCAGGTGCTGAGGCTGACATGAGAGACACAG 3417

DB 3361 GCCGATCAGCGGATCACCGAGTCTGCCAGAGTGTGAGAGCTGACATGAGAGACACAG 3420
QY 3418 GCTGAGATTTCTGCTCTGACAGAGGCTCTCAAGAGACAGAGCTGAAGGCTCGAGGCTC 3477
DB 3421 GCTGAGATTTCTGCTCTGACAGAGGCTCTCAAGAGACAGAGCTGAAGGCTCGAGGCTC 3480
QY 3478 TCTGACAGCTCAATGACCTGAGAGAGAGCATGTCTATGCTTGAATGAAATGCCCCAAGC 3537
DB 3481 TCTGACAGCTCAATGACCTGAGAGAGAGCATGTCTATGCTTGAATGAAATGCCCCAAGC 3540
QY 3538 TTAACAGCAGAGCTGAGAGCTGAACGAGAGCTCAACAGAGGCTTCTGAGAGAGCAAGCC 3597
DB 3541 TTAACAGCAGAGCTGAGAGCTGAACGAGAGCTCAACAGAGGCTTCTGAGAGAGCAAGCC 3600
QY 3598 AAAATTACAGCAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGACTCAAGACTG 3657
DB 3601 AAAATTACAGCAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGACTCAAGACTG 3660
QY 3658 CAAGAAGCTCTGATCGGCTGATCTACTGAAGACAGAAAGAGTGAAGTGAAGTATCAG 3717
DB 3661 CAAGAAGCTCTGATCGGCTGATCTACTGAAGACAGAAAGAGTGAAGTGAAGTATCAG 3720
QY 3718 CTGAAAAACATTCAGGTTCTTATTTCTCATGAAAAAGTGAATGGAAGGCACTATTCT 3777
DB 3721 CTGAAAAACATTCAGGTTCTTATTTCTCATGAAAAAGTGAATGGAAGGCACTATTCT 3780
QY 3778 CAACAAACCAACCTAATGATTTTCTGCAAGCCAAATGAGACCACTGTAAGAAAGAA 3837
DB 3781 CAACAAACCAACCTAATGATTTTCTGCAAGCCAAATGAGACCACTGTAAGAAAGAA 3840
QY 3838 AAGTTCTCTGCAATGAGCTGAAGCTGGGCTGGAAGAGGAAAGCTGCTGCT 3897
DB 3841 AAGTTCTCTGCAATGAGCTGAAGCTGGGCTGGAAGAGGAAAGCTGCTGCT 3900
QY 3898 GCGAGCTTGAAGAGAGCCCTTCAAGAACCCGATGAGCTCGGCTCCGCGGAGAGAA 3957
DB 3901 GCGAGCTTGAAGAGAGCCCTTCAAGAACCCGATGAGCTCGGCTCCGCGGAGAGAA 3960
QY 3958 GCTGCCACCCGAAAGCAACGAGACCAACCACTCATCGGCCAGCCGCGAGGCGAG 4017
DB 3961 GCTGCCACCCGAAAGCAACGAGACCAACCACTCATCGGCCAGCCGCGAGGCGAG 4020
QY 4018 CAGATCGCATGTCGCGCATGCGGCTGCGCAGAGACACAGCCAGTGGCCATGAGGCTG 4077
DB 4021 CAGATCGCATGTCGCGCATGCGGCTGCGCAGAGACACAGCCAGTGGCCATGAGGCTG 4080
QY 4078 CTGGCCCCGCAATCCAGCTCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTGGCGT 4137
DB 4081 CTGGCCCCGCAATCCAGCTCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTGGCGT 4140
QY 4138 CTTAAGAGAGCATGACACCAATATTTCTACCCGATTCACAGTGAAGCTGAACATGGCA 4197
DB 4141 CTTAAGAGAGCATGACACCAATATTTCTACCCGATTCACAGTGAAGCTGAACATGGCA 4200
QY 4198 GCCAACAAAGTGTGTGTGTCTGATACCGTGAACCTTGGACGCCAGGCAATCCAAATGT 4257
DB 4201 GCCAACAAAGTGTGTGTGTCTGATACCGTGAACCTTGGACGCCAGGCAATCCAAATGT 4260
QY 4258 CTGCAATGTCAAGT 4317
DB 4261 CTGCAATGTCAAGT 4320
QY 4318 TTGCTGTGTAAATTTGACACACACTTACCGAGGCTTCTGCGCGTGAACAAATGAATCC 4377
DB 4321 TTGCTGTGTAAATTTGACACACACTTACCGAGGCTTCTGCGCGTGAACAAATGAATCC 4380
QY 4378 CCAAGTCTCCAGACCAAGAGGCCAGAGAGAGCTTGAACCTGGAAGGCTGATGAAGGTG 4437
DB 4381 CCAAGTCTCCAGACCAAGAGGCCAGAGAGAGCTTGAACCTGGAAGGCTGATGAAGGTG 4440
QY 4438 CCAAGAAATTAACAAACAGAGACAGAGAGGCTGGGACAGAAAGTCAATTTGCTTGAAGGA 4497

Db 4441 CCCAGGAATTAACMAACGAGACAGAGAGCTGGAGACAGAGATGATCTGCTGGAGAG 4500
 QY 4498 TCAAAAGTCTCATTTATGACATGATGACGAGAGAGCTGGAGAGAGAGAGAGAG 4557
 Db 4501 TAAAAAGTCTCATTTATGACATGATGACGAGAGAGCTGGAGAGAGAGAGAGAG 4560
 QY 4558 TTTGAGCTGTGCTTCCGAGCGGAGATGATCTATTCATGATGCTGCTTCCGAA 4617
 Db 4561 TTTGAGCTGTGCTTCCGAGCGGAGATGATCTATTCATGATGCTGCTTCCGAA 4620
 QY 4618 CTGCGAATATACGCGCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 4677
 Db 4621 CTGCGAATATACGCGCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 4680
 QY 4678 ACCACCTGCTGGCCCGGAGAGAGAGCTGATGATGATGATGATGATGATGATGATGAT 4737
 Db 4681 ACCACCTGCTGGCCCGGAGAGAGAGCTGATGATGATGATGATGATGATGATGATGAT 4740
 QY 4738 GCTGAGGTACCGCTTGAATCAATGATGATGATGATGATGATGATGATGATGATGAT 4797
 Db 4741 GCTGAGGTACCGCTTGAATCAATGATGATGATGATGATGATGATGATGATGATGAT 4800
 QY 4798 GAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4857
 Db 4801 GAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4860
 QY 4858 GACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4917
 Db 4861 GACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4920
 QY 4918 CTCTAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4977
 Db 4921 CTCTAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4980
 QY 4978 TTTCAAAATTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5037
 Db 4981 TTTCAAAATTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5040
 QY 5038 GACAGTGTCTTGTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 5097
 Db 5041 GACAGTGTCTTGTGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 5100
 QY 5098 GCCCAGCCGACATCTCAACCCAACTTTTGAAGCTGCAAGGCTGCACTTGTGG 5157
 Db 5101 GCCCAGCCGACATCTCAACCCAACTTTTGAAGCTGCAAGGCTGCACTTGTGG 5160
 QY 5158 GCAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5217
 Db 5161 GCAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5220
 QY 5218 CTCCGCTCAAGGAAACCTTCAAGCAATCTGATCCGGAAGAGATGATGATGATGATGAT 5277
 Db 5221 CTCCGCTCAAGGAAACCTTCAAGCAATCTGATCCGGAAGAGATGATGATGATGATGAT 5280
 QY 5278 CCTGACAGCTGATCCTCACTTCAACCAATTACAGATCTCATGATGATGATGATGATGAT 5337
 Db 5281 CCTGACAGCTGATCCTCACTTCAACCAATTACAGATCTCATGATGATGATGATGATGAT 5340
 QY 5338 GAAATCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5397
 Db 5341 GAAATCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5400
 QY 5398 GCAAGT 5457
 Db 5401 GCAAGT 5460
 QY 5458 AGCGAG 5517
 Db 5461 AGCGAG 5520
 QY 5518 TCTTACGAG 5577
 Db 5521 TCTTACGAG 5580

QY 5578 GCCTACAGAGAAACCTATCTGTTTGTACCACTTCACTCACTGAGAGATGATGATGATGAT 5637
 Db 5581 GCCTACAGAGAAACCTATCTGTTTGTACCACTTCACTCACTGAGAGATGATGATGATGAT 5640
 QY 5638 CAG 5697
 Db 5641 CAG 5700
 QY 5698 TACCTGAG 5757
 Db 5701 TACCTGAG 5760
 QY 5758 TTAAG 5817
 Db 5761 TTAAG 5820
 QY 5818 GGCCTGCTCACTTCCGAG 5877
 Db 5821 GGCCTGCTCACTTCCGAG 5880
 QY 5878 ATCAG 5937
 Db 5881 ATCAG 5940
 QY 5938 CCAAG 5997
 Db 5941 CCAAG 6000
 QY 5998 GGCCTGCTCACTTCCGAG 6057
 Db 6001 GGCCTGCTCACTTCCGAG 6060
 QY 6058 TCCCGAG 6117
 Db 6061 TCCCGAG 6120
 QY 6118 ACCCGCTGCTTCCAG 6155
 Db 6121 ACCCGCTGCTTCCAG 6158

RESULT 7
 AB863436
 ID AB863436 standard; cDNA; 6189 BP.
 XX
 AC AB863436;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE RHO/RAC-interacting citron kinase-like human cDNA, designated NOV3b.
 XX
 KW Human; gene; ss; NOV; NOVX-associated disorder; cardiomyopathy;
 KW atherosclerosis; diabetes; cell signalling; metabolic pathway;
 KW cellular receptor; downstream effector; cancer; gene therapy;
 KW hypertension; congenital heart defect; aortic stenosis; obesity;
 KW infectious disease; anorexia; Alzheimer's disease; Parkinson's disease;
 KW neurodegenerative disorder; haemophilia; dyslipidemia; vaccine;
 KW haematopoietic disease; scleroderma; fertility; immunogen;
 KW idiopathic thrombocytopenic purpura; graft versus host disease;
 KW Crohn's disease; multiple sclerosis; cirrhosis; autoimmune disease;
 KW systemic lupus erythematosus; asthma; arthritis; psoriasis; allergy;
 KW stroke; anxiety; Lesch-Nyhan syndrome; schizophrenia; cerebellar ataxia;
 KW pain; alcoholism; transgenic.
 XX
 OS Homo sapiens.
 XX
 PN W0200226826-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 27-SEP-2001; 2001MC-US42336.
 XX

PR 27-SEP-2000; 2000US-235631P.
PR 27-SEP-2000; 2000US-235633P.
PR 27-SEP-2000; 2000US-235808P.
PR 27-SEP-2000; 2000US-236064P.
PR 27-SEP-2000; 2000US-236065P.
PR 27-SEP-2000; 2000US-236066P.
PR 28-SEP-2000; 2000US-236135P.
PR 03-OCT-2000; 2000US-237434P.
PR 05-OCT-2000; 2000US-238321P.
PR 06-OCT-2000; 2000US-238396P.
PR 06-OCT-2000; 2000US-238399P.
PR 16-MAR-2001; 2001US-276667P.
PR 31-MAY-2001; 2001US-294823P.
PR 12-JUL-2001; 2001US-304868P.
PR 26-SEP-2001; 2001US-0235631.
XX
XX (CURA-) CURAGEN CORP.
PI Gerlach VL, Macdougall JR, Smithson G, Millet I, Stone D,
PI Gunther E, Ellerman K, Grosse WM, Alsobrook JP, Lepley DM,
PI Burgess CB, Padigaru M, Kekuda R, Spytek KA, Leach MD;
PI Shimkets RA;
XX
XX WPI; 2002-499860/53.
DR P-PSDB; ABG78363.
XX

PT Novel isolated NOVX polypeptides and polynucleotides homologous to
PT attractin, plexin, papin-like family of proteins, useful for treating
PT atherosclerosis, diabetes, cancer, Alzheimer's disease, hemophilia and
PT stroke
XX
XX
XX Claim 8; Page 43-44; 308pp; English.

XX The invention discloses the isolated human polypeptides, and
XX polynucleotides encoding them, that have been designated NOVX. The
XX polypeptides, polynucleotides and antibodies are useful in treating or
XX preventing a NOVX-associated disorder which is cardiomyopathy,
XX atherosclerosis and diabetes in a human, where the disorder is related to
XX cell signal processing and metabolic pathway modulation. They can also be
XX used in determining the presence of, or predisposition to, a disease
XX associated with altered levels of the polypeptides and polynucleotides of
XX any one of the 13 sequences (NOV1-NOV8), for raising antibodies, for
XX identifying an agent that binds to, or that modulates the expression or
XX activity of the polypeptide, for identifying an agent which is cellular
XX receptor or downstream effector, for treating or preventing a
XX NOVX-associated disorder and as a pharmaceutical composition comprising
XX the polypeptide, polynucleotide or the antibody. The polypeptides and
XX polynucleotides are useful in diagnostic applications (e.g. as a marker
XX for cancerous cells or tissue types) where their amounts are assessed, or
XX for the manufacture of a medicament (e.g. gene therapy) for treating or
XX defects, aortic stenosis, obesity, infectious disease, anorexia, cancer,
XX Alzheimer's disease, Parkinson's disorder, neurodegenerative disorders,
XX haemophilia, dyslipidemias, haematopoietic diseases, scleroderma,
XX fertility, idiopathic thrombocytopenic purpura, graft versus host
XX diseases, Crohn's disease, multiple sclerosis, cirrhosis, autoimmune
XX disease, systemic lupus erythematosus, asthma, arthritis, psoriasis,
XX allergy, stroke, anxiety, Lesch-Nyman syndrome, schizophrenia,
XX cerebellar ataxia, pain and alcoholism. They may also be used as
XX immunogens to produce antibodies specific for the invention, and as
XX vaccines. Transgenic cells containing a NOVX expressing construct are
XX useful to produce non-human transgenic animals for studying the function
XX and/or activity of the NOVX proteins and for identifying and/or
XX evaluating modulators of NOVX protein activity. Transgenic cells
XX containing a NOVX expressing construct are useful to produce non-human
XX transgenic animals for studying the function and/or activity of the NOVX
XX proteins and for identifying and/or evaluating modulators of NOVX protein
XX activity. The sequences presented in ABS63431-ABS63444 are the human
XX NOV1-NOV8 CDNA's.

SO Sequence 6189 BP; 1742 A; 1554 C; 1690 G; 1203 T; 0 other;

Query Match

99.3%; Score 6121.8; DB 24; Length 6189;

Best Local Similarity 99.7%; Pred. No. 0;
Matches 6145; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY	1	ATGTTGAAGTTCAAAATATGAGCGCGGAATCTTTGGATGTGCTGCTGAACCAT	60
DB	1	ATGTTGAAGTTCAAAATATGAGCGCGGAATCTTTGGATGTGCTGCTGAACCAT	60
QY	61	GCCAGCGGCGCTCCAGAGTGAATCTGTTCTCCAGGGGAAACCACTTTATATGAC	120
DB	61	GCCAGCGGCGCTCCAGAGTGAATCTGTTCTCCAGGGGAAACCACTTTATATGAC	120
QY	121	CAGCAGATGTCCTCTTTCCCGAAGAGGATATTAAGTCCCTCTTGTCTCTTTGAA	180
DB	121	CAGCAGATGTCCTCTTTCCCGAAGAGGATATTAAGTCCCTCTTGTCTCTTTGAA	180
QY	181	GAATGACGTACAGCTGCTGATGATGAATTAAGCAGTGAACCTTTGTCGGAAGTAT	240
DB	181	GAATGACGTACAGCTGCTGATGATGAATTAAGCAGTGAACCTTTGTCGGAAGTAT	240
QY	241	TCCGACACCATAGCTGATTAAGAGAGCTCCAGCTTCGCGAAGAGCTTCGAGTCA	300
DB	241	TCCGACACCATAGCTGATTAAGAGAGCTCCAGCTTCGCGAAGAGCTTCGAGTCA	300
QY	301	AGTCTGTAGCTGTGTGCTCACTTTGCTGAAGTGAAGTGTAAAGAGAAACCGGG	360
DB	301	AGTCTGTAGCTGTGTGCTCACTTTGCTGAAGTGAAGTGTAAAGAGAAACCGGG	360
QY	361	GACATCATGCTATGAAGATGAAGAGAAAGAGCTTTATTTGGCCGAGAGCTTTCA	420
DB	361	GACATCATGCTATGAAGATGAAGAGAAAGAGCTTTATTTGGCCGAGAGCTTTCA	420
QY	421	TTTTTGAAGAGGCGGAACATATTAATCCGAAGCAAGACCCGTGATCCCAATTA	480
DB	421	TTTTTGAAGAGGCGGAACATATTAATCCGAAGCAAGACCCGTGATCCCAATTA	480
QY	481	CAGTATGCTTTTCAAGACAAAATCACTTTATCTGCTCAATGAATATCACTGAGGG	540
DB	481	CAGTATGCTTTTCAAGACAAAATCACTTTATCTGCTCAATGAATATCACTGAGGG	540
QY	541	GACTGTGTCATTTTGAATATGAATGAGACCACTTAATGAAAACTGATACAGTTT	600
DB	541	GACTGTGTCATTTTGAATATGAATGAGACCACTTAATGAAAACTGATACAGTTT	600
QY	601	TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCGTTCATGTGATGAGTACGATCGA	660
DB	601	TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCGTTCATGTGATGAGTACGATCGA	660
QY	661	GACATCAAGCTGAGAAATCTTCTGTTGACCGGACAGACATCAAGCTGTGGAATTT	720
DB	661	GACATCAAGCTGAGAAATCTTCTGTTGACCGGACAGACATCAAGCTGTGGAATTT	720
QY	721	GGATTCGCGCGGAAATGAATCAACAGATGATGATGCAAACTCCCGATTTGGAGC	780
DB	721	GGATTCGCGCGGAAATGAATCAACAGATGATGATGCAAACTCCCGATTTGGAGC	780
QY	781	CCAGATTAGATGCTCTTGAAGTCTGACTGTGATGAACGGGGATGAAAAAGCACCTAC	840
DB	781	CCAGATTAGATGCTCTTGAAGTCTGACTGTGATGAACGGGGATGAAAAAGCACCTAC	840
QY	841	GGCTGAGCTGATGCTGTGTGATGAGTGGCGCTGATTTCTTATGATGATTTATGAGAG	900
DB	841	GGCTGAGCTGATGCTGTGTGATGAGTGGCGCTGATTTCTTATGATGATTTATGAGAG	900
QY	901	TCCCGCTTGGCAGAGGAGAACTCTGCAACCTTCAATTAATTAATTTTCCAGCGG	960
DB	901	TCCCGCTTGGCAGAGGAGAACTCTGCAACCTTCAATTAATTAATTTTCCAGCGG	960
QY	961	TTTTTGAATTTTCAAGTACCCCAAGTGAAGTGAAGTCACTTTCTTATGATGATTTCAAG	1020
DB	961	TTTTTGAATTTTCAAGTACCCCAAGTGAAGTGAAGTCACTTTCTTATGATGATTTCAAG	1020
QY	1021	TTGTTGTGGCGGCAAAAGAGAGTGAAGTGAAGTCTTCTGCTGATCTTTCTTTC	1080
DB	1021	TTGTTGTGGCGGCAAAAGAGAGTGAAGTGAAGTCTTCTGCTGATCTTTCTTTC	1080

D 1018 TTGTTGCGGCGAAGAAAGAGAGACTGAATTGAAAGTCTTGTGCGCATCCTTCTTC 1077
Q 1081 TCTAAATGACTGGAAACAATTCGTAATCTCTCCCTCCCTGCTGCTCCACCCCAAG 1140
D 1078 TCTAAATGACTGGAAACAATTCGTAATCTCTCCCTCCCTGCTGCTCCACCCCAAG 1137
Q 1141 TCTGACGATGACACTTCGTAATTTTGTATGAAACAAGAAATTCGTGGTTTTCATCTCT 1200
D 1138 TCTGACGATGACACTTCGTAATTTTGTATGAAACAAGAAATTCGTGGTTTTCATCTCT 1197
Q 1201 CCGTGCAGCTGAGCCCTCAGAGCTTCTCGGTTGAAGAACTGCCGTTTGTGGGGTTTTCG 1260
D 1198 CCGTGCAGCTGAGCCCTCAGAGCTTCTCGGTTGAAGAACTGCCGTTTGTGGGGTTTTCG 1257
Q 1261 TACAGCAAGGCACTGGGGATTTCTTGTAGATCTGAGTCTGTGTGTCGAGTCTGACTCC 1320
D 1258 TACAGCAAGGCACTGGGGATTTCTTGTAGATCTGAGTCTGTGTGTCGAGTCTGACTCC 1317
Q 1321 CTTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAAAGACTTACAAAG 1380
D 1318 CTTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAAAGACTTACAAAG 1377
Q 1381 TCTCAGGCAAGTGTCAAGATGGAGCAGAAATGACCCGTTTATCATGCGAGAGTGTCA 1440
D 1378 TCTCAGGCAAGTGTCAAGATGGAGCAGAAATGACCCGTTTATCATGCGAGAGTGTCA 1437
Q 1441 GAGGTGAGGCTGTGCTTACTGTCAGAAAGAGTGGAGCTGAGGCTTCTGAGACTCAGAG 1500
D 1438 GAGGTGAGGCTGTGCTTACTGTCAGAAAGAGTGGAGCTGAGGCTTCTGAGACTCAGAG 1497
Q 1501 TCCCTCTGAGCAGAGACTTGTGCTACTCATCATCAAGATGCAATGCTTAAAGCGAAGT 1560
D 1498 TCCCTCTGAGCAGAGACTTGTGCTACTCATCATCAAGATGCAATGCTTAAAGCGAAGT 1557
Q 1561 TTGAGCAAGCAAGTGTGAGAGTGTCCAGAGAGATGACAAAGCACTGCACTTCTCAT 1620
D 1558 TTGAGCAAGCAAGTGTGAGAGTGTCCAGAGAGATGACAAAGCACTGCACTTCTCAT 1617
Q 1621 GATTCAGAGCAGAGCCGGAAGCTTCAGAAATCAAAAGCAGAGATCCAGGCTCAA 1680
D 1618 GATTCAGAGCAGAGCCGGAAGCTTCAGAAATCAAAAGCAGAGATCCAGGCTCAA 1677
Q 1681 GTGGAAGAAATGAGTGTGATGATGAATCAATGTTGAAGAGAGTCTTGTCTCAGCAAGAG 1740
D 1678 GTGGAAGAAATGAGTGTGATGATGAATCAATGTTGAAGAGAGTCTTGTCTCAGCAAGAG 1737
Q 1741 CGAGTATCTTACGATCTGAGCTGAGAGAGTCTCGGCTTGTCTGTAAGAAATTCAG 1800
D 1738 CGAGTATCTTACGATCTGAGCTGAGAGAGTCTCGGCTTGTCTGTAAGAAATTCAG 1797
Q 1801 CGGAAAGCAGAGATGTGAGCTTAACTGTGAAGCTTAAAGATCAAGGGAAGCTTGA 1860
D 1798 CGGAAAGCAGAGATGTGAGCTTAACTGTGAAGCTTAAAGATCAAGGGAAGCTTGA 1857
Q 1861 GTGGAAGAAATGAGTGTGAGAGATCAATGCTGAGCAGCAGCTCAAAATTCAGAG 1920
D 1858 GTGGAAGAAATGAGTGTGAGAGATCAATGCTGAGCAGCAGCTCAAAATTCAGAG 1917
Q 1921 CTTCAAGAGAACTGGAAGAGCTGTAAAGCCAGAGCCGAGCTGCTGCTGAG 1980
D 1918 CTTCAAGAGAACTGGAAGAGCTGTAAAGCCAGAGCCGAGCTGCTGCTGAG 1977
Q 1981 AATATCCGCGAAGGCAAGAGCCGAGAGGAGAGTGTGAAGCTGCAAGACCGAGAG 2040
D 1978 AATATCCGCGAAGGCAAGAGCCGAGAGGAGAGTGTGAAGAGCTGCAAGACCGAGAG 2037
Q 2041 GATTCTTCTGAAGGATCAGAAAGAGCTGTGAAGCTGAGAGCCGCTCATTTCTTG 2100
D 2038 GATTCTTCTGAAGGATCAGAAAGAGCTGTGAAGCTGAGAGCCGCTCATTTCTTG 2097
Q 2101 GAGAACAGGTAAGAGAGACTAGAGCAGTGAAGCTGAGAGAAAGCACTGAAGATGAG 2160
D 2098 GAGAACAGGTAAGAGAGACTAGAGCAGTGAAGCTGAGAGAAAGCACTGAAGATGAG 2157

Q 2161 ATCCAGCAAAATCCCAACAGATCCAGAGATGGCTGTAATAATTTCTGAGCTGCAAG 2220
D 2158 ATCCAGCAAAATCCCAACAGATCCAGAGATGGCTGTAATAATTTCTGAGCTGCAAG 2217
Q 2221 AAACATCGGAGGCCCAAGCTCTCAGCCGAGCACTTGAAGTGCATGAAACAGAAAG 2280
D 2218 AAACATCGGAGGCCCAAGCTCTCAGCCGAGCACTTGAAGTGCATGAAACAGAAAG 2277
Q 2281 CAGCACTTGTGAGAAATAATTAAGTGTGAATACTGATTAAGAAAGACTGCTGAC 2340
D 2278 CAGCACTTGTGAGAAATAATTAAGTGTGAATACTGATTAAGAAAGACTGCTGAC 2337
Q 2341 AAGAGACACTGAGAAACATGATGAGAGACAGCAGAGAGAGCCCATGAGAGGCAAA 2400
D 2338 AAGAGACACTGAGAAACATGATGAGAGACAGCAGAGAGAGCCCATGAGAGGCAAA 2397
Q 2401 ATTCTCAGGAAACAGAGCCGATGATCAATGCTATGATTTCCAGATCAGATCCTGAA 2460
D 2398 ATTCTCAGGAAACAGAGCCGATGATCAATGCTATGATTTCCAGATCAGATCCTGAA 2457
Q 2461 CAGAGATTTGTGAACTGTCTGAAGCCATTAACCTTGCAGCAATAGCAGCTTTTTC 2520
D 2458 CAGAGATTTGTGAACTGTCTGAAGCCATTAACCTTGCAGCAATAGCAGCTTTTTC 2517
Q 2521 CAAAGAAACATGAGAGGCCCAAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC 2580
D 2518 CAAAGAAACATGAGAGGCCCAAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC 2577
Q 2581 CTGAGAACACAGGCTGTGGAAGTGTGAAGCCCAAGAACTGAGAGACAGCTGGAG 2640
D 2578 CTGAGAACACAGGCTGTGGAAGTGTGAAGCCCAAGAACTGAGAGACAGCTGGAG 2637
Q 2641 AAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGTGGAATCTGAGAACAAATTTG 2700
D 2638 AAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGTGGAATCTGAGAACAAATTTG 2697
Q 2701 CGGAGGTCTAGTCTTGAAGACAGAGAGCAAGAACTGAGCTCAAGCCGACTCAGAG 2760
D 2698 CGGAGGTCTAGTCTTGAAGACAGAGAGCAAGAACTGAGCTCAAGCCGACTCAGAG 2757
Q 2761 CTACAGCTCTCCCTCAGAGAGGCGAGTCAAGATGAGCAGCCCTGCAAGGCTGCAAGGCG 2820
D 2758 CTACAGCTCTCCCTCAGAGAGGCGAGTCAAGATGAGCAGCCCTGCAAGGCTGCAAGGCG 2817
Q 2821 GCTCTGAGAGGCGAGCTTCCGCAAGGCGAAGCAAGAGCTGGAAGAGACCAAGCAGAGACT 2880
D 2818 GCTCTGAGAGGCGAGCTTCCGCAAGGCGAAGCAAGAGCTGGAAGAGACCAAGCAGAGACT 2877
Q 2881 GAAAGAGGATCCAGGCACTCAGGCACTAGAGATGAAATCCAGCGCAATTTGATGCT 2940
D 2878 GAAAGAGGATCCAGGCACTCAGGCACTAGAGATGAAATCCAGCGCAATTTGATGCT 2937
Q 2941 CTTCTTAACAGCTGATCTGTAATCAAGACCTGAGAGGAGCACTTAACCAAGCTGAGCCGAG 3000
D 2938 CTTCTTAACAGCTGATCTGTAATCAAGACCTGAGAGGAGCACTTAACCAAGCTGAGCCGAG 2997
Q 3001 GACAAAGCTGAACTCAACAAACAACTTCTTCTTCAAAACAACCTGAGAGGCTTCT 3060
D 2998 GACAAAGCTGAACTCAACAAACAACTTCTTCTTCAAAACAACCTGAGAGGCTTCT 3057
Q 3061 GAGCGCAAGCAGAGATGTGTAACAATGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3120
D 3058 GAGCGCAAGCAGAGATGTGTAACAATGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3117
Q 3121 ACAGAACAGAGATCAGCTTACAGACCAAGCAACAGATGAGAGCTTGAAGCAAG 3180
D 3118 ACAGAACAGAGATCAGCTTACAGACCAAGCAACAGATGAGAGCTTGAAGCAAG 3177
Q 3181 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGCCCTTAACAGATGAGCTGCTAGAA 3240
D 3178 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGCCCTTAACAGATGAGCTGCTAGAA 3237

QY	3241	AAAGAGCGGAGTGGGAGGCGCTTGGAGAGCGTCTCGGGTGATGAGAAAATCCCAAGTTTGAG	3300
Db	3238	AAAGAGCGGAGTGGGAGGCGCTTGGAGAGCGTCTCGGGTGATGAGAAAATCCCAAGTTTGAG	3297
QY	3301	TGTCGGGTTTGAGAGCTGCGAGAGAGAGTCTGGGACCCCGAGAAACAGAGCGGGCCAGAGCC	3360
Db	3298	TGTCGGGTTTGAGAGCTGCGAGAGAGAGTCTGGGACCCCGAGAAACAGAGCGGGCCAGAGCC	3357
QY	3361	GATCAGCGGATCACCCAGTCTCGCCAGGTGGTGGAGCTGGCCAGTGAAGAGCACAAAGCT	3420
Db	3358	GATCAGCGGATCACCCAGTCTCGCCAGGTGGTGGAGCTGGCCAGTGAAGAGCACAAAGCT	3417
QY	3421	GAAATTTCTGCTCTGCGAGAGCGCTCTCAAAAGCAGAACTGAAAGCCGAGAGCTCTCT	3480
Db	3418	GAAATTTCTGCTCTGCGAGAGCGCTCTCAAAAGCAGAACTGAAAGCCGAGAGCTCTCT	3477
QY	3481	GACAACCTCAATGACCTGGAGAGAAAGCATGCTATGCTTGAATGAAATGCCGAGCTTCA	3540
Db	3478	GACAACCTCAATGACCTGGAGAGAAAGCATGCTATGCTTGAATGAAATGCCGAGCTTCA	3537
QY	3541	CAGCAGAGCTGAGACTGMAAGAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCCAAA	3600
Db	3538	CAGCAGAGCTGAGACTGMAAGAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCCAAA	3597
QY	3601	TTACAGCAGCAGATGAGCCTCGAGAAAATCAATTTCGCTGACTCAAGACTGCAA	3660
Db	3598	TTACAGCAGCAGATGAGCCTCGAGAAAATCAATTTCGCTGACTCAAGACTGCAA	3657
QY	3661	GAAAGCTCTGAGCTGGGCTGATCTATCTGMAAGCAGAAAGAAATGACTTGGAGTACAGCTG	3720
Db	3658	GAAAGCTCTGAGCTGGGCTGATCTATCTGMAAGCAGAAAGAAATGACTTGGAGTACAGCTG	3717
QY	3721	GAAGAACATTCAAGTTCTGTATTTCTATGAAAAAGTGAATGGAAGCACTATTCTTCAA	3780
Db	3718	GAAGAACATTCAAGTTCTGTATTTCTATGAAAAAGTGAATGGAAGCACTATTCTTCAA	3777
QY	3781	CAAAACCAATCTATGATTTTCTGCAAGCCAAAATGACCAACTGCTAAAAAGAAAAG	3840
Db	3778	CAAAACCAATCTATGATTTTCTGCAAGCCAAAATGACCAACTGCTAAAAAGAAAAG	3837
QY	3841	GTTCCCTGCAAGTACAAATGAGGTGAGGTGGCCCTGGAGAAAGGAGAAAGCTCGTGTGCA	3900
Db	3838	GTTCCCTGCAAGTACAAATGAGGTGAGGTGGCCCTGGAGAAAGGAGAAAGCTCGTGTGCA	3897
QY	3901	GAGCTTAGAGAAAGCCCTTCAGAAAGCCCGCATCGAGCTCCGCGCGGAGAGAAAGCT	3960
Db	3898	GAGCTTAGAGAAAGCCCTTCAGAAAGCCCGCATCGAGCTCCGCGCGGAGAGAAAGCT	3957
QY	3961	GCCCAACCCGAAAGCAACGAGCCACCCATCTCACGCCAGCCACCGAGAGGACAGAG	4020
Db	3958	GCCCAACCCGAAAGCAACGAGCCACCCATCTCACGCCAGCCACCGAGAGGACAGAG	4017
QY	4021	ATCGCATGTCCGCGCATGTGGGGTCCCGAGAGACCAAGGCCAGTGGCCATGAGCTGTGG	4080
Db	4018	ATCGCATGTGTGCCATGTGGCGGTGGCGAGAGACCAAGGCCAGTGGCCATGAGCTGTGG	4077
QY	4081	GCCCCGCCATCAGCGCCGAGAAAGAGTCTTCAACTCAAGAGAAATTTAGTCGGGCTT	4140
Db	4078	GCCCCGCCATCAGCGCCGAGAAAGAGTCTTCAACTCAAGAGAAATTTAGTCGGGCTT	4137
QY	4141	AAGGAGCGCATGCAACCAATATTTCTCACCGATTCACGTAGAGACTGAAACATGCGAGCC	4200
Db	4138	AAGGAGCGCATGCAACCAATATTTCTCACCGATTCACGTAGAGACTGAAACATGCGAGCC	4197
QY	4201	ACAAAGTGTGTGTGTGTGTGATGAGAGTGCACATTGGAGATTCGAAATGCGAGCC	4260
Db	4198	ACAAAGTGTGTGTGTGTGTGATGAGAGTGCACATTGGAGATTCGAAATGCGAGCC	4257
QY	4261	GAATGTCAAGTGAATGTGTCAACCCCAAGTGTCTCACAGGCTTGGCCAGCCACTGGCGCTTG	4320
Db	4258	GAATGTCAAGTGAATGTGTCAACCCCAAGTGTCTCACAGGCTTGGCGGCTTG	4317
QY	4321	CTGTGTAATGTGCAACACTTCACCGAGGCTTCTGCCGTGACAAAATGAATCTCCCA	4380

Db	4318	CTGCTGAAATATGCCACAACATTCACCGAGCCTTCGCGTGACAAAATGAACCTCCCA	43777
Qy	4381	GGTCTCCAGACCAAGGAGCCAGCAGACGTCTGCACTTGAAAGGTGATGAAGTGGCC	44408
Db	4378	GGTCTCCAGACCAAGGAGCCAGCAGACGTCTGCACTTGAAAGGTGATGAAGTGGCC	44377
Qy	4441	AGGAATPAACAAACGAGGAGACGCAAGGCTGGGACAGAAAGTACATTGTCTGGAGGATCA	45000
Db	4438	AGGAATPAACAAACGAGGAGACGCAAGGCTGGGACAGAAAGTACATTGTCTGGAGGATCA	44977
Qy	4501	AAAGTCCATTTATATGACATGAAGCCAGAGAGCTGGACAGAGGCCGGTGGAAATTT	45666
Db	4498	AAAGTCCATTTATATGACATGAAGCCAGAGAGCTGGACAGAGGCCGGTGGAAATTT	45577
Qy	4561	GAGCTGTGCTTTCCCGACGGGAGTGTATCTATTCATGGTGCCTGTGTGCTCCGAATC	46208
Db	4558	GAGCTGTGCTTTCCCGACGGGAGTGTATCTATTCATGGTGCCTGTGTGCTCCGAATC	46177
Qy	4621	GCAATATACGACCAAGAGAGATGTGCCATACATACATGAAATGAAATCTCACCCGCAACC	46800
Db	4618	GCAATATACGACCAAGAGAGATGTGCCATACATACATGAAATGAAATCTCACCCGCAACC	46777
Qy	4681	ACCTGTGGCCCGGAGAGAACCTCTACTTGTGTAGCTCCAGCTTCCCTGACAAACAGGAC	47408
Db	4678	ACCTGTGGCCCGGAGAGAACCTCTACTTGTGTAGCTCCAGCTTCCCTGACAAACAGGAC	47377
Qy	4741	TGGATCACCGCCTTGAAATGCAATGTTGCCAGATGGAGAGTTTCTAGGAAAAACAGAA	48000
Db	4738	TGGATCACCGCCTTGAAATGCAATGTTGCCAGATGGAGAGTTTCTAGGAAAAACAGAA	47977
Qy	4801	GCTGATGCTAAACTCTTGAGAACTCCCTGCTGAAACGAGAAAGTATGACCGCTGTGAC	48600
Db	4798	GCTGATGCTAAACTCTTGAGAACTCCCTGCTGAAACGAGAAAGTATGACCGCTGTGAC	48577
Qy	4861	ATGAACCTGACGCTGCCCTTCAATGACACGGTGTGTTGTGGCACCGAGAGAGGCTC	49208
Db	4858	ATGAACCTGACGCTGCCCTTCAATGACACGGTGTGTTGTGGCACCGAGAGAGGCTC	49177
Qy	4921	TACGCCCTGAATGTGTTGAAAACTCCCTAACCATATGCCAGAAATTTGAGACATTTTC	49800
Db	4918	TACGCCCTGAATGTGTTGAAAACTCCCTAACCATATGCCAGAAATTTGAGACATTTTC	49777
Qy	4981	CAAAATTTATATATCAAGACCTGAGAGAGTACTCATGATAGAGGTGAAGACGGGCA	50400
Db	4978	CAAAATTTATATATCAAGACCTGAGAGAGTACTCATGATAGAGGTGAAGACGGGCA	50377
Qy	5041	CTGTGTCTTGAGAGGTGAAGAAAGTGAACAGTCCCTGGCCAGATCCCATCTGCTGCC	51000
Db	5038	CTGTGTCTTGAGAGGTGAAGAAAGTGAACAGTCCCTGGCCAGATCCCATCTGCTGCC	50977
Qy	5101	CAGCCCGACATCTACCCACATTTTGAAGCTGTCAAGGGCTGGCCACTTGTTGGGGGCA	51600
Db	5098	CAGCCCGACATCTACCCACATTTTGAAGCTGTCAAGGGCTGGCCACTTGTTGGGGGCA	51577
Qy	5161	GGCAAGATTGGAACGGGCTCTGCACTGTGACAGCCATGCCCCAGAAAGTGTCTATTTCT	52200
Db	5158	GGCAAGATTGGAACGGGCTCTGCACTGTGACAGCCATGCCCCAGAAAGTGTCTATTTCT	52177
Qy	5221	CGTATCAACGAAAACTCAGCAAAATATGCAATCCGGAAAAAGATTAAGACCTCAGAGCC	52800
Db	5218	CGTATCAACGAAAACTCAGCAAAATATGCAATCCGGAAAAAGATTAAGACCTCAGAGCC	52777
Qy	5281	TGCAGCTGTATCCACTTACCAATTAACAGTATCCTCATTTGGAACCAATAATTTCTAGAA	53400
Db	5278	TGCAGCTGTATCCACTTACCAATTAACAGTATCCTCATTTGGAACCAATAATTTCTAGAA	53377
Qy	5341	ATGCAACATGAAGACATGACAGCTCGAGGAATTTCTGGAATTAAGATGACATTTCTTGGCA	54000
Db	5338	ATGCAACATGAAGACATGACAGCTCGAGGAATTTCTGGAATTAAGATGACATTTCTTGGCA	53977
Qy	5401	CTGTGTGTGTTGCGGCTTTTCAACAGCTTCCCTGTCTCAATCGTGCAGGTGAAACAGC	54600

CC containing a NOVX expressing construct are useful to produce non-human
CC transgenic animals for studying the function and/or activity of the NOVX
CC proteins and for identifying and/or evaluating modulators of NOVX protein
CC activity. The sequences presented in AB563431-AB563444 are the human
CC NOV1-NOV8 cDNAs.

XX Sequence 6201 BP; 1736 A; 1552 C; 1707 G; 1206 T; 0 other;

Query Match 99.0%; Score 6105.4; DB 24; Length 6201;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 6131; Conservative 0; Mismatches 21; Indels 3; Gaps 1;

```
QY 1 ATGTTGAAGTTCAAAATATGAGCGCGGAATCCTTTGAGTCGTGTGCTGAACCCATT 60
DB 1 ATGTTGAAGTTCAAAATATGAGCGCGGAATCCTTTGAGTCGTGTGCTGAACCCATT 60
QY 61 GCCAGCGCGGCTCCGAGCGTGATCTGTTCTCCAGGGGAAACACCCCTTTAGACTCAA 120
DB 61 GCCAGCGCGGCTCCGAGCGTGATCTGTTCTCCAGGGGAAACACCCCTTTAGACTCAA 120
QY 121 CAGCAGATGTCCTCTTTCCGAGAAAGGATATTAGATGCCCTCTTTGTTCTCTTGAA 180
DB 121 CAGCAGATGTCCTCTTTCCGAGAAAGGATATTAGATGCCCTCTTTGTTCTCTTGAA 180
QY 181 GAATGCAATCAGCTGCTCTGATGAGATTAAGCAAGTGAACAATTGTCCGAAATAT 240
DB 181 GAATGCAATCAGCTGCTCTGATGAGATTAAGCAAGTGAACAATTGTCCGAAATAT 240
QY 241 TCCGACACCAATAGCTAGTTACAGAGAGCTCAGCCTTCGCGAAAGAGCTTCGAAATCAG 300
DB 241 TCCGACACCAATAGCTAGTTACAGAGAGCTCAGCCTTCGCGAAAGAGCTTCGAAATCAG 300
QY 301 AGTCTTGATAGTGTGTGATCACTTTGCTGAAGTGCAGGTGTGAAGAAAGCAACGGG 360
DB 301 AGTCTTGATAGTGTGTGATCACTTTGCTGAAGTGCAGGTGTGAAGAAAGCAACGGG 360
QY 361 GACATCTATGCTATGAAAGTGAAGAAAGAGGCTTTATTTGSCCCAGAGAGAGTTTCA 420
DB 361 GACATCTATGCTATGAAAGTGAAGAAAGAGGCTTTATTTGSCCCAGAGAGAGTTTCA 420
QY 421 TTTTGGAGAGAGGCGGAACATATTATCTCGAAGCAACAAGCCCGTGGATCCCCCAATTA 480
DB 421 TTTTGGAGAGAGGCGGAACATATTATCTCGAAGCAACAAGCCCGTGGATCCCCCAATTA 480
QY 481 CAGTATGCTTTTCAGAGCAAAAATCACTTTATCTGCTCATGGAATATCAGCTGAGAGG 540
DB 481 CAGTATGCTTTTCAGAGCAAAAATCACTTTATCTGCTCATGGAATATCAGCTGAGAGG 540
QY 541 GACTTGTCTGCTCTTTGAAATAGATATGAGAGCAAGTATGATAAAAAGCTGATACGTTT 600
DB 541 GACTTGTCTGCTCTTTGAAATAGATATGAGAGCAAGTATGATAAAAAGCTGATACGTTT 600
QY 601 TACTGAGTGAAGTATTTGCTGTTTCAAGGCTTCATCTGATGGGATACGTGATCCGA 660
DB 601 TACTGAGTGAAGTATTTGCTGTTTCAAGGCTTCATCTGATGGGATACGTGATCCGA 660
QY 661 GACATCAAGCCTTGAGAACTTCTGTTGAACCGCAAGACACATCAAGCTGTGATTTT 720
DB 661 GACATCAAGCCTTGAGAACTTCTGTTGAACCGCAAGACACATCAAGCTGTGATTTT 720
QY 721 GGATCGCGCGGAAATGAATTCAAAACAAGTGTGAATGSCCAAATCCCGATTTGGAGAC 780
DB 721 GGATCGCGCGGAAATGAATTCAAAACA--GATAATGSCCAAATCCCGATTTGGAGAC 780
QY 781 CCAGATTACATGAGCTCTGAAAGTCTGACTGTGATGAACGGGATGGAAGGACCTTAC 840
DB 781 CCAGATTACATGAGCTCTGAAAGTCTGACTGTGATGAACGGGATGGAAGGACCTTAC 840
QY 841 GGCCTGGAAGTCTGACTGTGAGTCAAGTGGCGGTGATTTGCTTATGAGATTTATGAGAGA 900
DB 841 GGCCTGGAAGTCTGACTGTGAGTCAAGTGGCGGTGATTTGCTTATGAGATTTATGAGAGA 900
QY 897 GGCCTGGAAGTCTGACTGTGAGTCAAGTGGCGGTGATTTGCTTATGAGATTTATGAGAGA 960
DB 897 GGCCTGGAAGTCTGACTGTGAGTCAAGTGGCGGTGATTTGCTTATGAGATTTATGAGAGA 960
QY 901 TCCCCCTTGCAGAGGAACTCTGCGAAGCCTTCAATTAACATTATGATTTCCAGCGG 960
DB 901 TCCCCCTTGCAGAGGAACTCTGCGAAGCCTTCAATTAACATTATGATTTCCAGCGG 960
```

```
DB 898 TCCCCCTTGCAGAGGAACTCTGCGAAGCCTTCAATTAACATTATGATTTCCAGCGG 957
QY 961 TTTTGAATTTTCCATATACCCCAAGAGAGAGACTTTTGTATCTGATTCGAAGC 1020
DB 961 TTTTGAATTTTCCATATACCCCAAGAGAGAGACTTTTGTATCTGATTCGAAGC 1020
QY 958 TTTTGAATTTTCCATATACCCCAAGAGAGAGACTTTTGTATCTGATTCGAAGC 1017
DB 958 TTTTGAATTTTCCATATACCCCAAGAGAGAGACTTTTGTATCTGATTCGAAGC 1017
QY 1021 TTGTTGTGCGGCGAGAAAGAGAGACTGAAGTTTGAAGTCTTTGTCGCCATCTTCTTC 1080
DB 1021 TTGTTGTGCGGCGAGAAAGAGAGACTGAAGTTTGAAGTCTTTGTCGCCATCTTCTTC 1080
QY 1018 TTGTTGTGCGGCGAGAAAGAGAGACTGAAGTTTGAAGTCTTTGTCGCCATCTTCTTC 1077
DB 1018 TTGTTGTGCGGCGAGAAAGAGAGACTGAAGTTTGAAGTCTTTGTCGCCATCTTCTTC 1077
QY 1081 TCTAAATTTGACTGGAACAACATTCCTGTCCTCCCTGTTGTTCCACCTCAAG 1140
DB 1081 TCTAAATTTGACTGGAACAACATTCCTGTCCTCCCTGTTGTTCCACCTCAAG 1140
QY 1078 TCTAAATTTGACTGGAACAACATTCCTGTCCTCCCTGTTGTTCCACCTCAAG 1137
DB 1078 TCTAAATTTGACTGGAACAACATTCCTGTCCTCCCTGTTGTTCCACCTCAAG 1137
QY 1141 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1200
DB 1141 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1200
QY 1138 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1197
DB 1138 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1197
QY 1201 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1260
DB 1201 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1260
QY 1198 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1257
DB 1198 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1257
QY 1261 TACAGCAAGGCACTGAGGATTTCTTGTGATCTGATCTGTTGTGTGGGTCTGACTCC 1320
DB 1261 TACAGCAAGGCACTGAGGATTTCTTGTGATCTGATCTGTTGTGTGGGTCTGACTCC 1320
QY 1258 TACAGCAAGGCACTGAGGATTTCTTGTGATCTGATCTGTTGTGTGGGTCTGACTCC 1317
DB 1258 TACAGCAAGGCACTGAGGATTTCTTGTGATCTGATCTGTTGTGTGGGTCTGACTCC 1317
QY 1321 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1380
DB 1321 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1380
QY 1318 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1377
DB 1318 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1377
QY 1381 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCGGTTTATCTCGAGAGTCTCA 1440
DB 1381 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCGGTTTATCTCGAGAGTCTCA 1440
QY 1378 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCGGTTTATCTCGAGAGTCTCA 1437
DB 1378 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCGGTTTATCTCGAGAGTCTCA 1437
QY 1441 GAGTGAAGGCTGTGCTTATGTCAGAAAGTGTGAAGTGAAGGCTCTGAGACTCAGAGA 1500
DB 1441 GAGTGAAGGCTGTGCTTATGTCAGAAAGTGTGAAGTGAAGGCTCTGAGACTCAGAGA 1500
QY 1438 GAGTGAAGGCTGTGCTTATGTCAGAAAGTGTGAAGTGAAGGCTCTGAGACTCAGAGA 1497
DB 1438 GAGTGAAGGCTGTGCTTATGTCAGAAAGTGTGAAGTGAAGGCTCTGAGACTCAGAGA 1497
QY 1501 TCCCTCTGAGAGAGGACCTTGCTACTCATCTCAAGAAATGCAATGCTTAAAGCCGAAGT 1560
DB 1501 TCCCTCTGAGAGAGGACCTTGCTACTCATCTCAAGAAATGCAATGCTTAAAGCCGAAGT 1560
QY 1498 TCCCTCTGAGAGAGGACCTTGCTACTCATCTCAAGAAATGCAATGCTTAAAGCCGAAGT 1557
DB 1498 TCCCTCTGAGAGAGGACCTTGCTACTCATCTCAAGAAATGCAATGCTTAAAGCCGAAGT 1557
QY 1561 TTGAGCAAGCAAGTGTGAGGAGTGTCCAGAGAGATGACAAACATCTGAGCTTCTCCAT 1620
DB 1561 TTGAGCAAGCAAGTGTGAGGAGTGTCCAGAGAGATGACAAACATCTGAGCTTCTCCAT 1620
QY 1558 TTGAGCAAGCAAGTGTGAGGAGTGTCCAGAGAGATGACAAACATCTGAGCTTCTCCAT 1617
DB 1558 TTGAGCAAGCAAGTGTGAGGAGTGTCCAGAGAGATGACAAACATCTGAGCTTCTCCAT 1617
QY 1621 GATATCAAGAGCAAGGCGGAAAGCTCCAAAGAAATCAAGAGCAGAGATCAAGGCTCAA 1680
DB 1621 GATATCAAGAGCAAGGCGGAAAGCTCCAAAGAAATCAAGAGCAGAGATCAAGGCTCAA 1680
QY 1618 GATATCAAGAGCAAGGCGGAAAGCTCCAAAGAAATCAAGAGCAGAGATCAAGGCTCAA 1677
DB 1618 GATATCAAGAGCAAGGCGGAAAGCTCCAAAGAAATCAAGAGCAGAGATCAAGGCTCAA 1677
QY 1681 GTGGAAGAAATGAGTGTGATGATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAGA 1740
DB 1681 GTGGAAGAAATGAGTGTGATGATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAGA 1740
QY 1678 GTGGAAGAAATGAGTGTGATGATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAGA 1737
DB 1678 GTGGAAGAAATGAGTGTGATGATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAGA 1737
QY 1741 CCGAGTATCTCTTCAAGATCTGAGTGAAGAGTCTCGGCTTCTGCTGAAGATTTCAAG 1800
DB 1741 CCGAGTATCTCTTCAAGATCTGAGTGAAGAGTCTCGGCTTCTGCTGAAGATTTCAAG 1800
QY 1738 CCGAGTATCTCTTCAAGATCTGAGTGAAGAGTCTCGGCTTCTGCTGAAGATTTCAAG 1797
DB 1738 CCGAGTATCTCTTCAAGATCTGAGTGAAGAGTCTCGGCTTCTGCTGAAGATTTCAAG 1797
QY 1801 CCGAAAGCGACAGAAATGTCAAGCATAACTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1860
DB 1801 CCGAAAGCGACAGAAATGTCAAGCATAACTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1860
QY 1798 CCGAAAGCGACAGAAATGTCAAGCATAACTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1857
DB 1798 CCGAAAGCGACAGAAATGTCAAGCATAACTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1857
QY 1861 GTGGAAGAAATATGCGAAACTGGAAGAAATCAATGTGAGAGAGAGCTCAAAATTCAGAG 1920
DB 1861 GTGGAAGAAATATGCGAAACTGGAAGAAATCAATGTGAGAGAGAGCTCAAAATTCAGAG 1920
QY 1858 GTGGAAGAAATATGCGAAACTGGAAGAAATCAATGTGAGAGAGAGCTCAAAATTCAGAG 1917
DB 1858 GTGGAAGAAATATGCGAAACTGGAAGAAATCAATGTGAGAGAGAGCTCAAAATTCAGAG 1917
QY 1921 TTCCAAAGAGAACTGGAAGAAAGCTGTAAAGCCAGCAAGAGCTCAAGGCTGTGAG 1980
DB 1921 TTCCAAAGAGAACTGGAAGAAAGCTGTAAAGCCAGCAAGAGCTCAAGGCTGTGAG 1980
QY 1918 TTCCAAAGAGAACTGGAAGAAAGCTGTAAAGCCAGCAAGAGCTGTGAG 1977
DB 1918 TTCCAAAGAGAACTGGAAGAAAGCTGTAAAGCCAGCAAGAGCTGTGAG 1977
QY 1981 AATATCCGCGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1981 AATATCCGCGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
```

Db	1978	AATATTCGGCAGGCAAAAGGAGCGAGCCGAGAGGGAGCTGGAGAAAGCTCAGAACCCGAGAG	2037
QY	2041	GATTCCTCTTGAAAGGCATCAGAAAAGAGCTGCTGGAAAGCTGAGGAAAGCGCGCATTCCTGTG	2100
Db	2038	GATTCCTCTGAAAGGCATCAGAAAAGAGCTGCTGGAAAGCTGAGGAAAGCGCGCATTCCTGTG	2097
QY	2101	GAGAACAGGTAAAGAGACTAGAGACCATGGACGCTAGAGAAAACAGACTAAGAGATGAC	2160
Db	2098	GAGAACAGGTAAAGAGACTAGAGACCATGGACGCTAGAGAAAACAGACTAAGAGATGAC	2157
QY	2161	ATCCACAACAAATCCCAACAGATCCAGACGATGGCTGATAAATCTGAGCTCGAAGAG	2220
Db	2158	ATCCACAACAAATCCCAACAGATCCAGACGATGGCTGATAAATCTGAGCTCGAAGAG	2217
QY	2221	AAACATCGGAGAGGCCCAAGTCTCAGGCCAGACCTTAGAAGTGCACCTGAAACAGAAAGAG	2280
Db	2218	AAACATCGGAGAGGCCCAAGTCTCAGGCCAGACCTTAGAAGTGCACCTGAAACAGAAAGAG	2277
QY	2281	CAGCACTATGAGAAAAAGATTAAATGTTTGGACATCATGATPAAAGAAAGACTTGACCTGAC	2340
Db	2278	CAGCACTATGAGAAAAAGATTAAATGTTTGGACATCATGATPAAAGAAAGACTTGACCTGAC	2337
QY	2341	AAGGAGACACTGGAGAAACATGATGACAGACACAGAGAGAGGCCCATAGAGAGGCGAAA	2400
Db	2338	AAGGAGACACTGGAGAAACATGATGACAGACACAGAGAGAGGCCCATAGAGAGGCGAAA	2397
QY	2401	ATTCTCAGCGAACAGAGAGCGATGATCATGCTATGATGATTCACAGATCAGATCCCTGGA	2460
Db	2398	ATTCTCAGCGAACAGAGAGCGATGATCATGCTATGATGATTCACAGATCAGATCCCTGGA	2457
QY	2461	CAAGAGATTGTGAACTGTCTGAGCCCAATAAATTGACAGCAAAATAGCATCTTTTACC	2520
Db	2458	CAAGAGATTGTGAACTGTCTGAGCCCAATAAATTGACAGCAAAATAGCATCTTTTACC	2517
QY	2521	CAAGAGAACATGAAGGCCCAAGAAAGATGATTTCTGAACCTCAGCGAACAGAAATTTTAC	2580
Db	2518	CAAGAGAACATGAAGGCCCAAGAAAGATGATTTCTGAACCTCAGCGAACAGAAATTTTAC	2577
QY	2581	CTGGAGACACAGAGCTGGGAGAGTTGGAGGCCCCAGAACCGAAAACTGGAGAGAGCATGGAG	2640
Db	2578	CTGGAGACACAGAGCTGGGAGAGTTGGAGGCCCCAGAACCGAAAACTGGAGAGAGCATGGAG	2637
QY	2641	AAAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGCTGGAACTGGAGACAAAGTTG	2700
Db	2638	AAAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGCTGGAACTGGAGACAAAGTTG	2697
QY	2701	CGGAGAGCTCAGCTCTAGAGACAGAGAGCAGAAACTGGAAGCTCAAGCGCCAGCTCACAAG	2760
Db	2698	CGGAGAGCTCAGCTCTAGAGACAGAGAGCAGAAACTGGAAGCTCAAGCGCCAGCTCACAAG	2757
QY	2761	CTACAGCTCTCCCTCGCAGAGACGCGAGATCAACAGTTGACAGCCCTGACAGGCTGACAGGGG	2820
Db	2758	CTACAGCTCTCCCTCGCAGAGAGCGAGATCAACAGTTGACAGCCCTGACAGGCTGACAGGGG	2817
QY	2821	GCCCTGGAGACCAGCTTCGCGCAGGCAAGACAGAGCTGGAAGAGACACAGCAGAGAGCT	2880
Db	2818	GCCCTGGAGACCAGCTTCGCGCAGGCAAGACAGAGCTGGAAGAGACACAGCAGAGAGCT	2877
QY	2881	GAAAGAGAGATCCAGGCACTCAGCGCACATAGAGATGAATCCAGCGCAAAATTTGATGCT	2940
Db	2878	GAAAGAGAGATCCAGGCACTCAGCGCACATAGAGATGAATCCAGCGCAAAATTTGATGCT	2937
QY	2941	CTTGATTAACAGCTGATCTGTAAATACAGACCTGGAGAGAGAGAGCTAAACCAAGCTGACCGAG	3000
Db	2938	CTTGATTAACAGCTGATCTGTAAATACAGACCTGGAGAGAGAGAGCTAAACCAAGCTGACCGAG	2997
QY	3001	GACAAACGCTGAACCTCAACCAACAAATTTTACTCTTGCCAAACMACTCGATGAGGCTTCT	3060
Db	2998	GACAAACGCTGAACCTCAACCAACAAATTTTACTCTTGCCAAACMACTCGATGAGGCTTCT	3057
QY	3061	GGCGCCAAACAGAGATTGTACAACTGCGAAAGTGAAGTGAACATCTCCGCGGAGATC	3120
Db	3058	GGCGCCAAACAGAGATTGTACAACTGCGAAAGTGAAGTGAACATCTCCGCGGAGATC	3117

QY	3121	ACGGAACGAGAGATGCACTTACCGCCGAAAGCAACGATGAGGCTCTGAAAGCACG	3188
Db	3118	ACGGAACGAGAGATGCACTTACCGCCGAAAGCAACGATGAGGCTCTGAAAGCACG	3177
QY	3181	TGCACCATGCTGGAGGAACAGGTCAATGAAATTTGGAGGCCCTTAACGATGAGCTGTTAGAA	3240
Db	3178	TGCACCATGCTGGAGGAACAGGTCAATGAAATTTGGAGGCCCTTAACGATGAGCTGTTAGAA	3237
QY	3241	AAAGACCGCAGTGGAGAGCCTGAGAGACGCTCCGCGGATGATGACAAATCCCACTTGAAG	3300
Db	3238	AAAGACCGCAGTGGAGAGCCTGAGAGAGCCTCCGCGGATGATGAGAAATCCCACTTTGAG	3299
QY	3301	TGTCGGGTTTGAGAGCTGCAGAGATGCTGGAACACCGAAGAACAGAGAGCGAGAGCC	3360
Db	3298	TGTCGGGTTTGAGAGCTGCAGAGATGCTGGAACACCGAAGAACAGAGAGCGAGAGCC	3355
QY	3361	GATCAGCGGATACCGAGCTCGCCAGAGGTGGAGCTGGCAGATGAAAGACCAAGGCT	3422
Db	3358	GATCAGCGGATACCGAGCTCGCCAGAGGTGGAGCTGGCAGATGAAAGACCAAGGCT	3417
QY	3421	GAGATTCTGCGCTCTGCAGAGGCTCTCAAGAGCAGAACTGAAGGCCGAGAGCTCTCT	3480
Db	3418	GAGATTCTGCGCTCTGCAGAGGCTCTCAAGAGCAGAACTGAAGGCCGAGAGCTCTCT	3477
QY	3481	GACAACTCAATGACCTGAGAGAGAGAGATGCTATGCTTGAATGAAATGCCCGAAGCTTA	3540
Db	3478	GACAACTCAATGACCTGAGAGAGAGAGATGCTATGCTTGAATGAAATGCCCGAAGCTTA	3537
QY	3541	CAGCAGAACTGGAGACTGAACGAGAGCTCAACAGAGGCTTCTGSAAGACCAAGCCAA	3600
Db	3538	CAGCAGAACTGGAGACTGAACGAGAGCTCAACAGAGGCTTCTGSAAGACCAAGCCAA	3597
QY	3601	TTACAGCAGCAGATGGAACCTGCAAAAAATCAATTTCCGCTGACTCAAGAGACTGCA	3660
Db	3598	TTACAGCAGCAGATGGAACCTGCAAAAAATCAATTTCCGCTGACTCAAGAGACTGCA	3657
QY	3661	GAGCTCAGAGTCGGGCTGATCTACTGTAAGCAGAAAGAGTGAATGAGATACACTG	3720
Db	3658	GAGCTCAGAGTCGGGCTGATCTACTGTAAGCAGAAAGAGTGAATGAGATACACTG	3717
QY	3721	GAATTCATTCAAGTTCTCTATTTCAATGAAGGTGAATATGAAAGCACTATTCTCAA	3780
Db	3718	GAATTCATTCAAGTTCTCTATTTCAATGAAGGTGAATATGAAAGCACTATTCTCAA	3777
QY	3781	CAAAACAATCATGTAATTTTCTGCAAGCCAAATGGAACCACTGCTAAAAAGAAAG	3840
Db	3778	CAAAACAATCATGTAATTTTCTGCAAGCCAAATGGAACCACTGCTAAAAAGAAAG	3837
QY	3841	GTTCCTCTGAGTACAAATGAGCTGAAGCTGGCCCTGGAGAGAGAAAGCTCGCTGCA	3900
Db	3838	GTTCCTCTGAGTACAAATGAGCTGAAGCTGGCCCTGGAGAGAGAAAGCTCGCTGCA	3897
QY	3901	GAGCTAAGAGAAAGCCCTTCAGAGAACCCGATGAGGCTCCGGTCCGCGCGAGAAAGCT	3960
Db	3898	GAGCTAAGAGAAAGCCCTTCAGAGAACCCGATGAGGCTCCGGTCCGCGCGAGAAAGCT	3957
QY	3961	GCCCAACGCAAGCAAGCAAGCAACCAACCACTCCAGCGCAAGCCAGCGGAGAGAGAG	4020
Db	3958	GCCCAACGCAAGCAAGCAAGCAACCAACCACTCCAGCGCAAGCCAGCGGAGAGAGAG	4017
QY	4021	ATCGCATGTCGCGCATCTGTCGGTCCGCAAGAGCAAGCCAGTGCATGAGCTGCTG	4080
Db	4018	ATCGCATGTCGCGCATCTGTCGGTCCGCAAGAGCAAGCCAGTGCATGAGCTGCTG	4077
QY	4081	GCCCGCATGCAAGCGGAGAAAGGATCTTCAATCCAGAGAAATTTAATCGGCGCTT	4140
Db	4078	GCCCGCATGCAAGCGGAGAAAGGATCTTCAATCCAGAGAAATTTAATCGGCGCTT	4137
QY	4141	AAGGAAGCATGACCAACAATATTTCTCAACCATTCACCTAGAGACTGAATGCAAGCC	4200
Db	4138	AAGGAAGCATGACCAACAATATTTCTCAACCATTCACCTAGAGACTGAATGCAAGCC	4197

QY 4201 ACAAGTGTGCTGTGTCTGTGATACCGTGAACCTTTGGACGCCAGGCATCCAAATGCTCTC 4260
DB 4198 ACAAGTGTGCTGTGTCTGTGATACCGTGAACCTTTGGACGCCAGGCATCCAAATGCTCTA 4257
QY 4261 GAATGTGAGTGAATGTGTCAACCCCAAGTGTCTCAACGGTCTTGGACGACCTCGGCTTG 4320
DB 4258 GAATGTGAGTGAATGTGTCAACCCCAAGTGTCTCAACGGTCTTGGACGACCTCGGCTTG 4317
QY 4321 CCTGTGAATATGCAACACATTCACCGAGGCTTCTGCGGTGACAAATGAATGAACTCCCA 4380
DB 4318 CCTGTGAATATGCAACACATTCACCGAGGCTTCTGCGGTGACAAATGAATGAACTCCCA 4377
QY 4381 GGTCTCCAGACCAAGGACCCAGAGAGCTTGCACCTGGAAAGGTGATGAAGTGC 4440
DB 4378 GGTCTCCAGACCAAGGACCCAGAGAGCTTGCACCTGGAAAGGTGATGAAGTGC 4437
QY 4441 AGGAATTAACAAAGAGAGACCAAGGCTGGAGACAGAAAGTAACTTGTCTCGAGGATCA 4500
DB 4438 AGGAATTAACAAAGAGAGACCAAGGCTGGAGACAGAAAGTAACTTGTCTCGAGGATCA 4497
QY 4501 AAAGTCTCATTTATGACAAATGAAGCCAGAGAGCTGACAGAGCGCGGTGAAGATTT 4560
DB 4498 AAAGTCTCATTTATGACAAATGAAGCCAGAGAGCTGACAGAGCGCGGTGAAGATTT 4557
QY 4561 GAGCTGCGCTTCCGACCGGGAGTGTATCTATCATGTGCGGTGTGTCTTCCGAATC 4620
DB 4558 GAGCTGCGCTTCCGACCGGGAGTGTATCTATCATGTGCGGTGTGTCTTCCGAATC 4617
QY 4621 GCATAATACAGCCAAAGACAGATGTCCCATACATCTGTAAGATGGAATCTCACCGCACAC 4680
DB 4618 GCATAATACAGCCAAAGACAGATGTCCCATACATCTGTAAGATGGAATCTCACCGCACAC 4677
QY 4681 ACCTGTGCGCGCGGAGAAACCTCTACTCTGACTAGCTCCAGCTTCCCTGACAAACAGCG 4740
DB 4678 ACCTGTGCGCGCGGAGAAACCTCTACTCTGACTAGCTCCAGCTTCCCTGACAAACAGCG 4737
QY 4741 TGGGTACACCGCTTAATCAATGTGTCCGAGGTGGAGAGTTCTTGGGAAAAAGCAGAA 4800
DB 4738 TGGGTACACCGCTTAATCAATGTGTGTGCGAGGTGGAGAGTTCTTGGGAAAAAGCAGAA 4797
QY 4801 GCTGATGCTAACTGCTTGGAAAACTCCCTGCTGAACCTGGAAGGTGATGACCGTCTAGAC 4860
DB 4798 GCTGATGCTAACTGCTTGGAAAACTCCCTGCTGAACCTGGAAGGTGATGACCGTCTAGAC 4857
QY 4861 ATGAATGCAACGCTGCTTCAAGTGAACAGGTGTGTGGTGGCAACCGAGAGAGGCTC 4920
DB 4858 ATGAATGCAACGCTGCTTCAAGTGAACAGGTGTGTGGTGGCAACCGAGAGAGGCTC 4917
QY 4921 TACGCCCTGAATGTCTGAAAAAATCCCTTAACCCATGTCCAGAGAAATGGAAGTCTTC 4980
DB 4918 TACGCCCTGAATGTCTGAAAAAATCCCTTAACCCATGTCCAGAGAAATGGAAGTCTTC 4977
QY 4981 CAAATTTATATTAATCAAGAGACCTGAGAACTACTCATGTATGACAGAGAAAGCGAGCA 5040
DB 4978 CAAATTTATATTAATCAAGAGACCTGAGAACTACTCATGTATGACAGAGAAAGCGAGCA 5037
QY 5041 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGGCCCACTGCTGCTG 5100
DB 5038 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGGCCCACTGCTGCTG 5097
QY 5101 CAGCCGACATCTCAACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTTTGGGGCA 5160
DB 5098 CAGCCGACATCTCAACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTTTGGGGCA 5157
QY 5161 GGCAGAGATTGAGAACGGGCTTGTGATCTGTGACGACATCCAGCAAAATGCTCATTTCTC 5220
DB 5158 GGCAGAGATTGAGAACGGGCTTGTGATCTGTGACGACATCCAGCAAAATGCTCATTTCTC 5217
QY 5221 CGCTACAGAGAAACCTGACAAATTAAGTATCCGAAAGATTAAGATTAAGATTAAGATTAAG 5280
DB 5218 CGCTACAGAGAAACCTGACAAATTAAGTATCCGAAAGATTAAGATTAAGATTAAGATTAAG 5277
QY 5281 TGCAGCTGTATCCACTTCAACCAATTACAGTATCTCATGTGAACCAATTAATTCAAGAA 5340

DB 5278 TGCAGCTGTATCCACTTCAACCAATTACAGTATCTCATGTGAACCAATTAATTCAAGAA 5337
QY 5341 ATGCAATGAAGAGTACACGCTGAGAGAAATTCCTGATTAAGATTAAGATTAAGATTAAG 5400
DB 5338 ATGCAATGAAGAGTACACGCTGAGAGAAATTCCTGATTAAGATTAAGATTAAGATTAAG 5397
QY 5401 CCTGTGTGTTTCCCGCTCTTCAACAGCTTCCCTGTCTCATTCGTCAGGTGCAAGC 5460
DB 5398 CCTGTGTGTTTCCCGCTCTTCAACAGCTTCCCTGTCTCATTCGTCAGGTGCAAGC 5457
QY 5461 GCAGGACAGAGAGAGTACTTGTGTGTTTCCAGAAATTTGAGAGTGTGAGATTC 5520
DB 5458 GCAGGACAGAGAGAGTACTTGTGTGTTTCCAGAAATTTGAGAGTGTGAGATTC 5517
QY 5521 TACGAAAGCTTACCGGACAGACATCTCAAGTGAAGTGTACTTGTGCTTGGCTTGGC 5580
DB 5518 TACGAAAGCTTACCGGACAGACATCTCAAGTGAAGTGTACTTGTGCTTGGCTTGGC 5577
QY 5581 TACAGAGAACCCATCTGTTGTGACCCACTTCACTCACTGCAAGTAAATTGAGATCCAG 5640
DB 5578 TACAGAGAACCCATCTGTTGTGACCCACTTCACTCACTGCAAGTAAATTGAGATCCAG 5637
QY 5641 GCACGCTCTTCAAGAGAGACCCCTGCGAGCGTACTGACATCCCGAAACCGCGCTAC 5700
DB 5638 GCACGCTCTTCAAGAGAGACCCCTGCGAGCGTACTGACATCCCGAAACCGCGCTAC 5697
QY 5701 CTGGGCGCTGCCATTTCTCAAGAGAGATTTACTTGGCGTCTCATACAGATTAATTA 5760
DB 5698 CTGGGCGCTGCCATTTCTCAAGAGAGATTTACTTGGCGTCTCATACAGATTAATTA 5757
QY 5761 AGGTCATTTGCTGCAAGAGAAACCTGTGAAGAGTTCGGGCACTGAACCAACCGGGGCG 5820
DB 5758 AGGTCATTTGCTGCAAGAGAAACCTGTGAAGAGTTCGGGCACTGAACCAACCGGGGCG 5817
QY 5821 CCGTCACTTCCGACAGAGCCCAACAGAGAGCCCAACCACTGACAAAGAGCATC 5880
DB 5818 CCGTCACTTCCGACAGAGCCCAACAGAGAGCCCAACCACTGACAAAGAGCATC 5877
QY 5881 ACCAAGCGGTGGCTTCCAGAGCCCAAGCGCGGCCGGAAGGCCCAAGCCCAAGGCCA 5940
DB 5878 ACCAAGCGGTGGCTTCCAGAGCCCAAGCGCGGCCGGAAGGCCCAAGCCCAAGGCCA 5937
QY 5941 AGCACAACCCCAACGCTAAGCGAGAGAGGAGGAGCCGAGCTGCGAGAGCAAGTCTCTGCG 6000
DB 5938 AGCACAACCCCAACGCTAAGCGAGAGAGGAGGAGGAGCCGAGCTGCGAGAGCAAGTCTCTGCG 5997
QY 6001 CCGCCCTTGAAGCAGAGAAATCCCGCGAGTCTCAGACAACGAGAGAGAGCGATCC 6060
DB 5998 CCGCCCTTGAAGCAGAGAAATCCCGCGAGTCTCAGACAACGAGAGAGAGCGATCC 6057
QY 6061 CCGGAGAGCTTGTGAAGACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6120
DB 6058 CCGGAGAGCTTGTGAAGACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6117
QY 6121 CCGTGTCCCAAGTGAACAGGTCTGGGACCACTC 6155
DB 6118 CCGTGTCCCAAGTGAACAGGTGAAGCAAGTTC 6152

RESULT 9
ABQ78871
ID ABQ78871 standard; cDNA; 5877 BP.
XX
XX ABQ78871;
AC
AC 10-OCT-2002 (first entry)
DT
DT
XX
XX Human kinase cDNA #2.
DE
DE Human; kinase; serine-threonine kinase; nootropic; cytotstatic;
KW Human; kinase; serine-threonine kinase; gene therapy; mental disorder; cancer;
KW gene; ss.

XX OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..5877
XX /tag= a
XX /product= "Kinase"
XX MO200259325-A2.
XX 01-AUG-2002.
XX 20-DEC-2001; 2001WO-US50497.
XX 27-DEC-2000; 2000US-258335P.
XX (LEXI-) LEXICON GENETICS INC.
XX Yu X, Miranda M, Friiddle CJ;
XX WPI; 2002-599796/64.
XX P-PSDB; ABB81928.
XX Novel polynucleotide encoding human proteins that are structurally
XX similar to animal kinases, useful for drug screening, diagnosis, in
XX gene therapy of disorders and diseases e.g. cancer and pharmacogenomic
XX applications
XX
XX Disclosure; Page 44-45; 50pp; English.
XX
XX The invention relates to a novel human protein that shares structural
XX similarity with animal kinases, including serine-threonine kinases,
XX particularly Cdkron rho-interacting kinases. The proteins of the
XX invention have nootropic and cytoskeletal activity. The polynucleotides may
XX have a use in gene therapy. The encoded novel polypeptides are useful for
XX generating antibodies, as reagents in diagnostic assays, for identifying
XX other cellular gene products related to NHP and as reagents in assays for
XX screening for compounds that are useful in the treatment of mental,
XX biological or medical disorders and diseases including cancer. The
XX sequence encodes a novel human kinase of the invention.
XX
XX Sequence 5877 BP; 1680 A; 1433 C; 1583 G; 1181 T; 0 other;
XX
XX Query Match 94.7%; Score 5837; DB 24; Length 5877;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 5837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGTTGAAGTCAAAATATGAGCGCGAATCCTTTGATGCTGCTGAACCAT 60
XX 1 ATGTTGAAGTCAAAATATGAGCGCGAATCCTTTGATGCTGCTGAACCAT 60
XX 61 GCCAGCGCGGCTCCAGGCTGAATCTGTTCTTCAAGGGAACCACTTTATGACTCA 120
XX 61 GCCAGCGCGGCTCCAGGCTGAATCTGTTCTTCAAGGGAACCACTTTATGACTCA 120
XX 121 CAGCAGATGCTCTCTTTCCCGAGAGGATATTAGATGCCCTTTGTTCTTTGAA 180
XX 121 CAGCAGATGCTCTCTTTCCCGAGAGGATATTAGATGCCCTTTGTTCTTTGAA 180
XX 121 CAGCAGATGCTCTCTTTCCCGAGAGGATATTAGATGCCCTTTGTTCTTTGAA 180
XX 181 GAATGAGTCAAGCTCTGATGAGATTAAGCAGCTGAGCAACTTTGTCGGAAGTAT 240
XX 181 GAATGAGTCAAGCTCTGATGAGATTAAGCAGCTGAGCAACTTTGTCGGAAGTAT 240
XX 181 GAATGAGTCAAGCTCTGATGAGATTAAGCAGCTGAGCAACTTTGTCGGAAGTAT 240
XX 241 TCCGACACCATAGCTGATTAAGAGAGCTCAGCTTCGGAAGGCTTCGAGTCA 300
XX 241 TCCGACACCATAGCTGATTAAGAGAGCTCAGCTTCGGAAGGCTTCGAGTCA 300
XX 301 AGTCTTAGAGTGTGTGCTCTTGTGTAAGTGAAGTGTGTAAGAGGAACCGAG 360
XX 301 AGTCTTAGAGTGTGTGCTCTTGTGTAAGTGAAGTGTGTAAGAGGAACCGAG 360
XX 361 GACATCTATGCTATGAAAGTGAAGAAGAGCTTTATGGCCGAGAGCGATTTC 420
XX 361 GACATCTATGCTATGAAAGTGAAGAAGAGCTTTATGGCCGAGAGCGATTTC 420

Db 361 GACATCTATGCTATGAAAGTGAAGAAGAGCTTTATGGCCGAGAGCGATTTC 420
Qy 421 TTTTGGAGAGAGCGAACAATATATCTGAGACAGAGCCGTGATCCCCAATTA 480
Db 421 TTTTGGAGAGAGCGAACAATATATCTGAGACAGAGCCGTGATCCCCAATTA 480
Qy 481 CAGTATGCTCTTCAAGAGCAAAATCACTTTATCTGATCATGGAATATCAGCTGAGG 540
Db 481 CAGTATGCTCTTCAAGAGCAAAATCACTTTATCTGATCATGGAATATCAGCTGAGG 540
Qy 541 GACTTGTCTCATCTTTGATATGATATGAGAGCAAGTATGAAACCTGATACAGTT 600
Db 541 GACTTGTCTCATCTTTGATATGATATGAGAGCAAGTATGAAACCTGATACAGTT 600
Qy 601 TACCTAGCTGAGCTGATTTTGGCTGTTCACAGAGCTTCACTGATGGAATGATGATCA 660
Db 601 TACCTAGCTGAGCTGATTTTGGCTGTTCACAGAGCTTCACTGATGGAATGATGATCA 660
Qy 661 GACATCAAGCTGAGAGCAATCTCTGTTGACCGCACAGGACATCAAGCTGTGATTTT 720
Db 661 GACATCAAGCTGAGAGCAATCTCTGTTGACCGCACAGGACATCAAGCTGTGATTTT 720
Qy 721 GATCTGCGCGCAAAATGATTAACAAGATGATGATGCAAACTCCGATTTGGAGCC 780
Db 721 GATCTGCGCGCAAAATGATTAACAAGATGATGATGCAAACTCCGATTTGGAGCC 780
Qy 781 CCAGATTACATGCTCTGAAAGTCTGATGATGAAACGAGGATGAAAGCACTTAC 840
Db 781 CCAGATTACATGCTCTGAAAGTCTGATGATGAAACGAGGATGAAAGCACTTAC 840
Qy 841 GGCCTGAGCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 GGCCTGAGCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Qy 901 TCCCTTGGAGAGAGGAACTCTGCGCAAGCTTCAATTAATGATGATGATGATGATGAT 960
Db 901 TCCCTTGGAGAGAGGAACTCTGCGCAAGCTTCAATTAATGATGATGATGATGATGAT 960
Qy 961 TTTTGAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 961 TTTTGAATTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Qy 1021 TTTTGGAGAGAGGAACTCTGCGCAAGCTTCAATTAATGATGATGATGATGATGATGAT 1080
Db 1021 TTTTGGAGAGAGGAACTCTGCGCAAGCTTCAATTAATGATGATGATGATGATGATGAT 1080
Qy 1081 TCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 TCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Qy 1141 TCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1141 TCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy 1201 CCGTGCAGTGAAGCCCTCAGGCTTCCGAGTGAAGAACTGCGTTGTTGGGTTTTCG 1260
Db 1201 CCGTGCAGTGAAGCCCTCAGGCTTCCGAGTGAAGAACTGCGTTGTTGGGTTTTCG 1260
Qy 1261 TACAGAGAGCACTGAGGATTTCTGTGATGATGATGATGATGATGATGATGATGAT 1320
Db 1261 TACAGAGAGCACTGAGGATTTCTGTGATGATGATGATGATGATGATGATGATGAT 1320
Qy 1321 CCGTGCAGTGAAGCCCTCAGGCTTCCGAGTGAAGAACTGCGTTGTTGGGTTTTCG 1380
Db 1321 CCGTGCAGTGAAGCCCTCAGGCTTCCGAGTGAAGAACTGCGTTGTTGGGTTTTCG 1380
Qy 1381 TCTGAGAGCAAGTGTCAAGATGAGAGAGAACTCCGCTTACATGAGAGTGTCA 1440
Db 1381 TCTGAGAGCAAGTGTCAAGATGAGAGAGAACTCCGCTTACATGAGAGTGTCA 1440
Qy 1441 GAGGTGAGAGCTGTGCTTATGATGAGAGAGGATGAGAGCTTCAAGAGCTCAAGAG 1500
Db 1441 GAGGTGAGAGCTGTGCTTATGATGAGAGAGGATGAGAGCTTCAAGAGCTCAAGAG 1500

```
QY 1501 TCCCTCTGAGAGGACTTGTCTCACTCAATGACAGATGACAGTAAAGCAGT 1560
DB 1501 TCCCTCTGAGAGGAGCTTGTCTACCTACATCAAAATGACAGTAAAGCAGT 1560
QY 1561 TTGGAGAGCAGCAGGATGAGAGGTGTCCAGAGAGATGACAAAGCACTGCGCTTCTCAT 1620
DB 1561 TTGGAGAGCAGCAGGATGAGAGGTGTCCAGAGAGATGACAAAGCACTGCGCTTCTCAT 1620
QY 1621 GATATCAGAGAGCAGAGCCCGAAGCTCCAGAAATCAAAGAGCAGAGTCCAGGCTCAA 1680
DB 1621 GATATCAGAGAGCAGAGCCCGAAGCTCCAGAAATCAAAGAGCAGAGTCCAGGCTCAA 1680
QY 1681 GTGGAGAAATGAGGTGATGATGAATCACTTGAAGAGGATCTTGTCTCAGCAAGAGA 1740
DB 1681 GTGGAGAAATGAGGTGATGATGAATCACTTGAAGAGGATCTTGTCTCAGCAAGAGA 1740
QY 1741 CGGAGGATCTCTACAAATCTGAGCTGAGAGAGTCTCGCTTGTCTGTGAAGATTCCAG 1800
DB 1741 CGGAGGATCTCTACAAATCTGAGAGAGTCTCGCTTGTCTGTGAAGATTCCAG 1800
QY 1801 CGGAAAGCAGCAGAAATGTGACATTAACCTGTTGAAGGCTAAAGATCAAGGAGCCTGAA 1860
DB 1801 CGGAAAGCAGCAGAAATGTGACATTAACCTGTTGAAGGCTAAAGATCAAGGAGCCTGAA 1860
QY 1861 GTGGAGAGATATGCGGAACTGGAAGATCAATGCTGAGAGCAGCTCAAAATTCAGAG 1920
DB 1861 GTGGAGAGATATGCGGAACTGGAAGATCAATGCTGAGAGCAGCTCAAAATTCAGAG 1920
QY 1921 CTCAGAGAGAACTGAGAGAGGCTGTAAAGCCAGAGCCAGAGCCGCCAGCTGCTGCAAG 1980
DB 1921 CTCAGAGAGAACTGAGAGAGGCTGTAAAGCCAGAGCCAGAGCCGCCAGCTGCTGCAAG 1980
QY 1981 AATATCCGCCAGCAGAAAGAGCCAGAGCCAGAGAGCTGAGAGAGCTGAGAAACCGAGAG 2040
DB 1981 AATATCCGCCAGCAGAAAGAGCCAGAGCCAGAGAGCTGAGAGAGCTGAGAAACCGAGAG 2040
QY 2041 GATTCCTCTGAGAGGCTCAAGAAAGAGGCTGTGAAAGCTGAGAGAGCCGCCATTTCTGT 2100
DB 2041 GATTCCTCTGAGAGGCTCAAGAAAGAGGCTGTGAAAGCTGAGAGAGCCGCCATTTCTGT 2100
QY 2101 GAGAACAGAGTAAAGAGCTAGAGACCATGAGCCGTAGAGAAACAGACTGAAGAGTGA 2160
DB 2101 GAGAACAGAGTAAAGAGCTAGAGACCATGAGCCGTAGAGAAACAGACTGAAGAGTGA 2160
QY 2161 ATCCAGACAAAAATCCCAACAGATCCAGCAGATGCTGATTAATTTCTGAGCTCCAGAGAG 2220
DB 2161 ATCCAGACAAAAATCCCAACAGATCCAGCAGATGCTGATTAATTTCTGAGCTCCAGAGAG 2220
QY 2221 AAAACATCGGAGAGCCCAAGCTCTCAGCCCAAGCACTTAAGAGTCACTGAAACAGAAAGAG 2280
DB 2221 AAAACATCGGAGAGCCCAAGCTCTCAGCCCAAGCACTTAAGAGTCACTGAAACAGAAAGAG 2280
QY 2281 CAGCACTATGAGAGAAAGATTAAGTGTGGAACAATCAGATTAAGAAAGAAACCTGCGTGA 2340
DB 2281 CAGCACTATGAGAGAAAGATTAAGTGTGGAACAATCAGATTAAGAAAGAAACCTGCGTGA 2340
QY 2341 AAAGAGACACTGAGAAACATGATGACAGACACAGAGAGAGGCCCATGAGAAAGGCAAA 2400
DB 2341 AAAGAGACACTGAGAAACATGATGACAGACACAGAGAGAGGCCCATGAGAAAGGCAAA 2400
QY 2401 ATTCTCAGAGAACAGAGGCGATGATCAATGCTATGATTTCAAGATCAGATCCCTGGA 2460
DB 2401 ATTCTCAGAGAACAGAGGCGATGATCAATGCTATGATTTCAAGATCAGATCCCTGGA 2460
QY 2461 CAGAGGATTTGGAACCTGTCTGAAGCCAAATACTTGCAGCAATAGCAGTCTTTTTC 2520
DB 2461 CAGAGGATTTGGAACCTGTCTGAAGCCAAATACTTGCAGCAATAGCAGTCTTTTTC 2520
QY 2521 CAAGAGAACATGAAAGGCCCAAGAGATGATTTTGAATTCAGGCAACAGAAATTTTAC 2580
DB 2521 CAAGAGAACATGAAAGGCCCAAGAGATGATTTTGAATTCAGGCAACAGAAATTTTAC 2580
```

```
QY 2581 CTGAGACACAGGCTGAGAAAGTTGAGAGCCCAAGAACGAAAACTGAGAGAGCAGCTGAG 2640
DB 2581 CTGAGACACAGGCTGAGAAAGTTGAGAGCCCAAGAACGAAAACTGAGAGAGCAGCTGAG 2640
QY 2641 AAGATCACCCACCAAGACCACTGACAAAGATCGGCTGTGAACTGAGACAAAGATTG 2700
DB 2641 AAGATCACCCACCAAGACCACTGACAAAGATCGGCTGTGAACTGAGACAAAGATTG 2700
QY 2701 CGGAGAGTCACTTAAAGCACAGAGAGCAGAAACTGAGACTCAAGCCCACTCAGAG 2760
DB 2701 CGGAGAGTCACTTAAAGCACAGAGAGCAGAAACTGAGACTCAAGCCCACTCAGAG 2760
QY 2761 CTACAGCTCTCCCTGAGAGAGCCGAGTCAAGTTGACAGCCCTGACAGGCTGACCGGCG 2820
DB 2761 CTACAGCTCTCCCTGAGAGAGCCGAGTCAAGTTGACAGCCCTGACAGGCTGACCGGCG 2820
QY 2821 GCCCTGAGAGCCAGCTTGCAGGCGGAAGACAGAGCTGAGAGACCAAGCAGAGCT 2880
DB 2821 GCCCTGAGAGCCAGCTTGCAGGCGGAAGACAGAGCTGAGAGACCAAGCAGAGCT 2880
QY 2881 GAAGAGAGATCCAGGCACTCAGGCACTAAGAGATGAATCCAGGCCAAATTTGATGCT 2940
DB 2881 GAAGAGAGATCCAGGCACTCAGGCACTAAGAGATGAATCCAGGCCAAATTTGATGCT 2940
QY 2941 CTTCTGAAGAGCTGTACTGTAAATCAGAGACCTGAGAGAGCAGCTAAACAGCTGACCGAG 3000
DB 2941 CTTCTGAAGAGCTGTACTGTAAATCAGAGACCTGAGAGAGCAGCTAAACAGCTGACCGAG 3000
QY 3001 GACACGCTGAATCTCAACACCAAACTTCTTCTCAACCAACTGATGAGGCTTCT 3060
DB 3001 GACACGCTGAATCTCAACACCAAACTTCTTCTCAACCAACTGATGAGGCTTCT 3060
QY 3061 GCGGCCACAGACAGATTTGTAACAATGCAAGTGAAGTGAACCATCTCCGCCGAGATC 3120
DB 3061 GCGGCCACAGACAGATTTGTAACAATGCAAGTGAAGTGAACCATCTCCGCCGAGATC 3120
QY 3121 ACGGAAACAGAGATGAGCTTACAGCCAGAAACCAAGATGAGAGCTTGAAGACACAG 3180
DB 3121 ACGGAAACAGAGATGAGCTTACAGCCAGAAACCAAGATGAGAGCTTGAAGACACAG 3180
QY 3181 TGCAACATGCTGAGAGAACAGATCATGATTTGAGAGCCCTTAAACGATGAGCTGTAGAA 3240
DB 3181 TGCAACATGCTGAGAGAACAGATCATGATTTGAGAGCCCTTAAACGATGAGCTGTAGAA 3240
QY 3241 AAAAGCCGCAATGAGAGGCTCTGAGAGAGGCTCTGAGTGAAGAGAAATCCCAATTTGAG 3300
DB 3241 AAAAGCCGCAATGAGAGGCTCTGAGAGAGGCTCTGAGTGAAGAGAAATCCCAATTTGAG 3300
QY 3301 TGTGCGGTTTCAAGAGCTGAGAGATGCTGACACCGAGAAACAGAGCAGGCGAGAGCC 3360
DB 3301 TGTGCGGTTTCAAGAGCTGAGAGATGCTGACACCGAGAAACAGAGCAGGCGAGAGCC 3360
QY 3361 GATCAGCGGATACCGAGCTCTGCGCAGAGGTGAGACTGAGCAAGAGCT 3420
DB 3361 GATCAGCGGATACCGAGCTCTGCGCAGAGGTGAGACTGAGCAAGAGCT 3420
QY 3421 GAGATTTCTGCTCTCAGAGAGGCTCTCAAAGAGCAGAAAGCTGAAGGCGAGGCTCTCT 3480
DB 3421 GAGATTTCTGCTCTCAGAGAGGCTCTCAAAGAGCAGAAAGCTGAAGGCGAGGCTCTCT 3480
QY 3481 GACAAAGCTCAATGACCTGAGAGAGAGATGCTATGCTTGAATGAATGCCCCGAAGCTTA 3540
DB 3481 GACAAAGCTCAATGACCTGAGAGAGAGATGCTATGCTTGAATGAATGCCCCGAAGCTTA 3540
QY 3541 CAGCAGAAAGCTGAGAGACTGAACAGAGCTCAACAGAGGCTTCTGAAAGAGCAAGCCAAA 3600
DB 3541 CAGCAGAAAGCTGAGAGACTGAACAGAGCTCAACAGAGGCTTCTGAAAGAGCAAGCCAAA 3600
QY 3601 TTACAGCAGAGATGAGCCTGCAAGAAATACATTTTCCGCTGACCTCAAGGACTGGA 3660
DB 3601 TTACAGCAGAGATGAGCCTGCAAGAAATACATTTTCCGCTGACCTCAAGGACTGGA 3660
QY 3661 GAAGCTTAGATCGGCGTGAATCTAGTAAGACAGAAAGAGTGAAGCTTGAAGATTCAGCTG 3720
```

Db 3661 GAAGCTTAGATCGGGCTGATCTACTGAGACAGAAAGATGACTTGGATTACAGCTTG 3720
Qy 3721 GAAACATTCAGGTTCTCTATTCTCATGCAAAAAGGTGAAATGGAAGGCACTATTTCTCAA 3780
Db 3721 GAAACATTCAGGTTCTCTATTCTCATGCAAAAAGGTGAAATGGAAGGCACTATTTCTCAA 3780
Qy 3781 CAACACCAACATTCGATTTTCTGCAAGCCCAATATGACCACTGCTTAAAAAGAAAAG 3840
Db 3781 CAACACCAACATTCGATTTTCTGCAAGCCCAATATGACCACTGCTTAAAAAGAAAAG 3840
Qy 3841 GTTCTCTGCAAGTACAATAGCTGAGCTGAGCCTCTGAGAAAGAGAAAGCTGCTGTCGA 3900
Db 3841 GTTCTCTGCAAGTACAATAGCTGAGCTGAGCCTCTGAGAAAGAGAAAGCTGCTGTCGA 3900
Qy 3901 GAGCTTAGAGAGAGCCCTTTCAGAGAGACCCGCAATGCAAGCTCCGCTCCGGAGAGAGCT 3960
Db 3901 GAGCTTAGAGAGAGCCCTTTCAGAGAGACCCGCAATGCAAGCTCCGCTCCGGAGAGAGCT 3960
Qy 3961 GCCCACCAGAAAG 4020
Db 3961 GCCCACCAGAAAG 4020
Qy 4021 ATGCGCATGTCCGCGCATGTGCGGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4021 ATGCGCATGTCCGCGCATGTGCGGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Qy 4081 GCGCCCGCATCCAGCCGAG 4140
Db 4081 GCGCCCGCATCCAGCCGAG 4140
Qy 4141 AAGGAAAGCATGAGACCAAAATATTTCTCAACCGATTCAAGTGAAGTGAAGTGAAGTGAAG 4200
Db 4141 AAGGAAAGCATGAGACCAAAATATTTCTCAACCGATTCAAGTGAAGTGAAGTGAAGTGAAG 4200
Qy 4201 ACAAGT 4260
Db 4201 ACAAGT 4260
Qy 4261 GAATGTCAAGT 4320
Db 4261 GAATGTCAAGT 4320
Qy 4321 CCTGCTGAATATGACCACTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380
Db 4321 CCTGCTGAATATGACCACTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380
Qy 4381 GGTCTTCAGACCAAG 4440
Db 4381 GGTCTTCAGACCAAG 4440
Qy 4441 AAGGAATTAACAAAG 4500
Db 4441 AAGGAATTAACAAAG 4500
Qy 4501 AAAAGTCTCATTTATGACAAATGAAAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4560
Db 4501 AAAAGTCTCATTTATGACAAATGAAAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4560
Qy 4561 GAGCTGTGCTTCCGAG 4620
Db 4561 GAGCTGTGCTTCCGAG 4620
Qy 4621 GCAATTAACAGCCAAAG 4680
Db 4621 GCAATTAACAGCCAAAG 4680
Qy 4681 AAGCTGTGAG 4740
Db 4681 AAGCTGTGAG 4740
Qy 4741 TGGGTACCGGCTTGAATCAAGTGTGCAAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800
Db 4741 TGGGTACCGGCTTGAATCAAGTGTGCAAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800

Db 4741 TGGGTACCGGCTTGAATCAAGTGTGCAAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800
Qy 4801 GGTGATGTAAACGCTTTGAGAACTCCCTGCTGAAACGAGAGAGAGAGAGAGAGAGAGAGAG 4860
Db 4801 GGTGATGTAAACGCTTTGAGAACTCCCTGCTGAAACGAGAGAGAGAGAGAGAGAGAGAGAG 4860
Qy 4861 ATGAACGTGACGCTGCCCTTCAAGTGAACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Db 4861 ATGAACGTGACGCTGCCCTTCAAGTGAACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 4921 TACGCCCTGAATGTCTTGAAGAACTCCCTTAACCATATGCTCCAGAGATTTGAGAGAGCTTC 4980
Db 4921 TACGCCCTGAATGTCTTGAAGAACTCCCTTAACCATATGCTCCAGAGATTTGAGAGAGCTTC 4980
Qy 4981 CAATTTATTTATTAACAAG 5040
Db 4981 CAATTTATTTATTAACAAG 5040
Qy 5041 CTGTGTCTTGTGAG 5100
Db 5041 CTGTGTCTTGTGAG 5100
Qy 5101 CAGCCGCAATCTCAACCAATTTTGAAGCTGTCAAGGCTGTCAAGGCTGTCCACTTGTGGGGCA 5160
Db 5101 CAGCCGCAATCTCAACCAATTTTGAAGCTGTCAAGGCTGTCAAGGCTGTCCACTTGTGGGGCA 5160
Qy 5161 GGCAGATTTGAG 5220
Db 5161 GGCAGATTTGAG 5220
Qy 5221 CGCTTACAG 5280
Db 5221 CGCTTACAG 5280
Qy 5281 TGGAGCTGTATCCACTTACCAATTAACGATCTCTAATTTGAGAGAGAGAGAGAGAGAGAGAG 5340
Db 5281 TGGAGCTGTATCCACTTACCAATTAACGATCTCTAATTTGAGAGAGAGAGAGAGAGAGAGAG 5340
Qy 5341 ATGACATGAAG 5400
Db 5341 ATGACATGAAG 5400
Qy 5401 CCTGTGTGTGTGCGCCTCTTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5460
Db 5401 CCTGTGTGTGTGCGCCTCTTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5460
Qy 5461 GCAAGGAG 5520
Db 5461 GCAAGGAG 5520
Qy 5521 TACGAG 5580
Db 5521 TACGAG 5580
Qy 5581 TACAG 5640
Db 5581 TACAG 5640
Qy 5641 GCAAGCTCTTCAAG 5700
Db 5641 GCAAGCTCTTCAAG 5700
Qy 5701 CTGGGAG 5760
Db 5701 CTGGGAG 5760
Qy 5761 AAGGTCATTTGTGCAAG 5820
Db 5761 AAGGTCATTTGTGCAAG 5820
Qy 5821 CCGTCCACTCTCCGAG 5837
Db 5821 CCGTCCACTCTCCGAG 5837

RESULT 10
 ID AAD39191 standard; cDNA; 6574 BP.
 AC AAD39191;
 DT 04-OCT-2002 (first entry)
 XX Human MDPK cDNA.
 DE Human; myotonic dystrophy type protein kinase; MDPK; 13245 protein;
 KW tumorigenesis; tumour growth; tumour metastasis; viral infection;
 KW skeletal muscle disorder; muscular dystrophy; myotonic dystrophy;
 KW immune disorder; neoplastic disorder; gene therapy; gene; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..18
 FT CDS /tag= a
 FT /tag= b
 FT /product= "Human MDPK protein"
 FT /tag= c
 FT /note= "This region is specifically referred
 in claim 1 as SEQ ID NO:3"
 FT 3'UTR 6181..6574
 FT /tag= d
 PN WO00234896-A2.
 PD 02-MAY-2002.
 PF 23-OCT-2001; 2001WO-US50636.
 PR 23-OCT-2000; 2000US-242429P.
 PA (MILL-) MILLENNIUM PHARM INC.
 XX Kapeller-Libermann R;
 DR WPI; 2002-479720/51.
 DR P-Psdb; AAE24079.
 PT Human myotonic dystrophy type protein kinase polypeptide and
 PT inhibiting tumorigenesis, tumor growth, tumor metastasis and viral
 PT infection
 PS Claim 1; Fig 1; 148pp; English.
 XX The invention relates to human myotonic dystrophy type protein kinase
 CC (MDPK) polypeptides designated as 13245 and nucleic acid molecules
 CC encoding such polypeptides. 13245 molecules are used to develop
 CC preventing, inhibiting, alleviating or curing MDPK-related disorders.
 CC Polypeptides of the invention are used to develop MDPK-related disorders.
 CC therapeutic agents for 13245-mediated or related disorders such as
 CC tumorigenesis, tumour growth, tumour metastasis, viral infection
 CC of a cell, skeletal muscle disorders (e.g. muscular and myotonic
 CC dystrophies), immune disorders and neoplastic disorders. The
 CC invention is also used in gene therapy. The present sequence is
 CC human MDPK cDNA.
 XX
 SQ Sequence 6574 BP; 1877 A; 1611 C; 1776 G; 1310 T; 0 other;
 Query Match 91.8%; Score 5661.4; DB 24; Length 6574;
 Best Local Similarity 95.8%; Pred. No. 0;
 Matches 5944; Conservative 0; Mismatches 16; Indels 243; Gaps 3;

QY 1 ATGTGAAGTTCATAATATGAGCGGAAATCTTTGATGCTGTGTAACCAT 60
 Db 19 ATGTGAAGTTCATAATATGAGCGGAAATCTTTGATGCTGTGTAACCAT 78
 QY 61 GCCAGCCGGGCTCCAGGCTGAATCTTTCTCCAGGAGAAACACCTTTATGACTCA 120
 Db 79 GCCAGCCGGGCTCCAGGCTGAATCTTTCTCCAGGAGAAACACCTTTATGACTCA 138
 QY 121 CAGCAGATGCTCTCTTCCCGAGAGAGGATTTGATGCTCTTTGCTCTTTGAA 180
 Db 139 CAGCAGATGCTCTCTTCCCGAGAGAGGATTTGATGCTCTTTGCTCTTTGAA 198
 QY 181 GAATGAGTCAAGCTCTCTCTGATGAAGATTTAAGCAGTGAACATTTGTCGGAAGTAT 240
 Db 199 GAATGAGTCAAGCTCTCTGATGAAGATTTAAGCAGTGAACATTTGTCGGAAGTAT 258
 QY 241 TCCGACACCATATGCTGAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 300
 Db 259 TCCGACACCATATGCTGAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 318
 QY 301 AGCTTTGATGCTGTGTGCTCACTTGTGATGAGTGTGAAGAGAGAAAGCAACCGG 360
 Db 319 AGCTTTGATGCTGTGTGCTCACTTGTGATGAGTGTGAAGAGAGAAAGCAACCGG 378
 QY 361 GACATCTATGCTATGAAGATGATGAAGAAAGCTTTATTTGGCCAGAGAGGTTTCA 420
 Db 379 GACATCTATGCTATGAAGATGATGAAGAAAGCTTTATTTGGCCAGAGAGGTTTCA 438
 QY 421 TTTTGTGAGAAAGACCGGACATATTATCTCGAAGCAACGCGGTGATCCCGCAATTA 480
 Db 439 TTTTGTGAGAAAGACCGGACATATTATCTCGAAGCAACGCGGTGATCCCGCAATTA 498
 QY 481 CAGTATGCTTTGAGAAACCAATCATCTTATCTGTGATGAGATTAAGCTTGAAGG 540
 Db 499 CAGTATGCTTTGAGAAACCAATCATCTTATCTGTGATGAGATTAAGCTTGAAGG 558
 QY 541 GACTTGTCTCACTTTTGAATGATGAGACCAATTATGATGAAGAAACCTGATACGTT 600
 Db 559 GACTTGTCTCACTTTTGAATGATGAGACCAATTATGATGAAGAAACCTGATACGTT 618
 QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCTTCACTTGTGATGAGATGAGTCA 660
 Db 619 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCTTCACTTGTGATGAGATGAGTCA 678
 QY 661 GACATCAAGCTGAGAAACATCTGTGACCGGACAGGACATCAACCTGCTGATTT 720
 Db 679 GACATCAAGCTGAGAAACATCTGTGACCGGACAGGACATCAACCTGCTGATTT 738
 QY 721 GATCTGCGGAAATGATTAATCAACAGATGATGATGCAACCTGCTGATTT 780
 Db 739 GATCTGCGGAAATGATTAATCAACAGATGATGATGCAACCTGCTGATTT 798
 QY 781 CCAGATTAAGCTGAGCTGCTGAGTGTGATGAAACCGGAGTGAAGAAAGCACTTAC 840
 Db 799 CCAGATTAAGCTGAGCTGCTGAGTGTGATGAAACCGGAGTGAAGAAAGCACTTAC 858
 QY 841 GGCTGAGCTGAGCTGCTGAGTGTGATGAAACCGGAGTGAAGAAAGCACTTAC 900
 Db 859 GGCTGAGCTGAGCTGCTGAGTGTGATGAAACCGGAGTGAAGAAAGCACTTAC 918
 QY 901 TCCCTCTTGGAGAGGAACTCTGCAAGAACTTCAATTAATTAATTAATTAATTAAT 960
 Db 919 TCCCTCTTGGAGAGGAACTCTGCAAGAACTTCAATTAATTAATTAATTAATTAAT 978
 QY 961 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAATTTGATGATTTCAAGG 1020
 Db 979 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAATTTGATGATTTCAAGG 1038
 QY 1021 TTGTGTGCGGCGAGAAAGAGAGTGAATTTGAAGTCTTTGCTGCAATCTTTCTTC 1080
 Db 1039 TTGTGTGCGGCGAGAAAGAGAGTGAATTTGAAGTCTTTGCTGCAATCTTTCTTC 1098
 QY 1081 TCTAAATTTGATGAGAAACATTTGTAATCTTCTCTCTCTCTCTCTCTCTCTCTCT 1140

Db 1099 TCTAATTTGATGGAACAACATTCGTAACTCTCCCTCCCTCGTTCCACCTCTAG 1158
Qy 1141 TCTGCGATGACACTCTCCAAATTTTGTATACCAAGAAAGATTTGTGGTTTCACTCT 1200
Db 1159 TCTGCGATGACACTCTCCAAATTTTGTATACCAAGAAAGATTTGTGGTTTCACTCT 1218
Qy 1201 CCGTCCAGCTGAGCCCTCAGGCTTCTCGGATGAAAGATCCGCTTGTGGGTTTTCG 1260
Db 1219 CCGTCCAGCTGAGCCCTCAGGCTTCTCGGATGAAAGATCCGCTTGTGGGTTTTCG 1278
Qy 1261 TACAGCAAGGACCTGGGATTTCTTGTATGATCTGATCTGTTGTGCGGCTTGGACTCC 1320
Db 1279 TACAGCAAGGACCTGGGATTTCTTGTATGATCTGATCTGTTGTGCGGCTTGGACTCC 1338
Qy 1321 CCTGCAAGACTAGTCTCATGAAAAGAACTTCTCATCAAAACCAAGGCTTCAAGAC 1380
Db 1339 CCTGCAAGACTAGTCTCATGAAAAGAACTTCTCATCAAAACCAAGGCTTCAAGAC 1398
Qy 1381 TCTCAGGACAAGTGTCAACAGATGAGCAGGAAATGACCCGCTTACATCGGAGATGTCA 1440
Db 1399 TCTCAGGACAAGTGTCAACAGATGAGCAGGAAATGACCCGCTTACATCGGAGATGTCA 1458
Qy 1441 GAGGTGAGGCTGTGCTTATGACAGAGAGGTGAGAGGCTTGAAGCTTGAAGCTTGAAG 1500
Db 1459 GAGGTGAGGCTGTGCTTATGACAGAGAGGTGAGAGGCTTGAAGCTTGAAGCTTGAAG 1518
Qy 1501 TCCCTCTGAGCAGGACCTTGTCTACCTACATCAAGATGACAGTACTTAAAGCAGAT 1560
Db 1519 TCCCTCTGAGCAGGACCTTGTCTACCTACATCAAGATGACAGTACTTAAAGCAGAT 1578
Qy 1561 TTGAGCAGACGAGCTGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620
Db 1579 TTGAGCAGACGAGCTGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1638
Qy 1621 GATATCAGAGCAGACGCGGAGAGCTCAAGAAATCAAGAGCAGAGATACAGGCTCAA 1680
Db 1639 GATATCAGAGCAGACGCGGAGAGCTCAAGAAATCAAGAGCAGAGATACAGGCTCAA 1698
Qy 1661 GTGAGAAATAGAGTTGATGATGATCAATGTTGAAGAGATCTTGTCTCAGCAAGAGA 1740
Db 1699 GTGAGAAATAGAGTTGATGATGATCAATGTTGAAGAGATCTTGTCTCAGCAAGAGA 1758
Qy 1741 CGGAGTGTCTCTGAAATCTGAGCTGAGAGTCTGGGCTTGTGCTGAAGAAATTCAG 1800
Db 1759 CGGAGTGTCTCTGAAATCTGAGCTGAGAGTCTGGGCTTGTGCTGAAGAAATTCAG 1818
Qy 1801 CGGAAAGCAGAGATGTCAAGATTAATCTGTTGAAGCTTGAAGATCAAGGAGAGCTGAA 1860
Db 1819 CGGAAAGCAGAGATGTCAAGATTAATCTGTTGAAGCTTGAAGATCAAGGAGAGCTGAA 1878
Qy 1861 GTGAGAGAAATAGAGAACTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1920
Db 1879 GTGAGAGAAATAGAGAACTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1938
Qy 1921 CTCCAAGAGAACTGAGAGAGGCTGTAAAGCCAGCAAGAGGACCAAGCTGCTGAG 1980
Db 1939 CTCCAAGAGAACTGAGAGAGGCT- 1962
Qy 1961 AATATCCGCGCAGGCAAGAGAGCGGAGCGAGGAGGCTGAGAAAGCTGCAAGACCGAG 2040
Db 1963 -----GAAAGAGAGCGAGCGGAGGAGGAGGCTGAGAAAGCTGCAAGACCGAG 2010
Qy 2041 GATTCCTTGAAGGATCAGAAAGAGAGTGTGAGAGCTGAGAAAGCGCGCATCTCTG 2100
Db 2011 GATTCCTTGAAGGATCAGAAAGAGAGTGTGAGAGCTGAGAAAGCGCGCATCTCTG 2070
Qy 2101 GAGAAACAAGGTAAAGAGCTAGAGACCATGAGCGGTAGAGAAACAGACTGAAGATGAC 2160
Db 2071 GAGAAACAAGGTAAAGAGCTAGAGACCATGAGCGGTAGAGAAACAGACTGAAGATGAC 2130
Qy 2161 ATCCAGACAAGATCCCAACAGATCCAGAGATGCTGATAAATTTTGGAGCTCGAAGAG 2220
Db 2131 ATCCAGACAAGATCCCAACAGATCCAGAGATGCTGATAAATTTTGGAGCTCGAAGAG 2190
Qy 2221 AAACATCGGAGAGCCCAAGATCTCAGAGCCAGACCTTAAAGATGACCTGAAAGAGAG 2280
Db 2191 AAACATCGGAGAGCCCAAGATCTCAGAGCCAGACCTTAAAGATGACCTGAAAGAGAG 2250
Qy 2281 CAGCACTATGAGGAAAGATTAAGTTTGAAGTGTGACCAATGAGATTAAGAAAGACTGCTGAC 2340
Db 2251 CAGCACTATGAGGAAAGATTAAGTTTGAAGTGTGACCAATGAGATTAAGAAAGACTGCTGAC 2310
Qy 2341 AAGAGACACTGAGAGAACTGATGACAGACAGAGAGAGGCGCCATGAGAGAGGCAAA 2400
Db 2311 AAGAGACACTGAGAGAACTGATGACAGACAGAGAGAGGCGCCATGAGAGAGGCAAA 2370
Qy 2401 ATTCCTAGGAAACAGAGGCGGATGATCAATGCTATGAGATTCAGAGTCAATCTCTGAA 2460
Db 2371 ATTCCTAGGAAACAGAGGCGGATGATCAATGCTATGAGATTCAGAGTCAATCTCTGAA 2430
Qy 2461 CAGAGATTTGGAACCTGTCTGAAGCAATTAATCTGACCAATAGCAGTCTTTTACC 2520
Db 2431 CAGAGATTTGGAACCTGTCTGAAGCAATTAATCTGACCAATAGCAGTCTTTTACC 2490
Qy 2521 CAAAGGAAATGAAAGGCGCCAAAGAGAGATTTTCTGAATCTGAGCAACAGAAATTTAC 2580
Db 2491 CAAAGGAAATGAAAGGCGCCAAAGAGAGATTTTCTGAATCTGAGCAACAGAAATTTAC 2550
Qy 2581 CTGAGACACAGGCTTGGAGTTGAGAGCCCAAGAACCGGAACTGAGAGAGCAGCTGAG 2640
Db 2551 CTGAGACACAGGCTTGGAGTTGAGAGCCCAAGAACCGGAACTGAGAGAGCAGCTGAG 2610
Qy 2641 AAGATGAGCCACCAAGACCAAGTGAACAAATTCGCTGAGAACTGAGAGCAAGATTG 2700
Db 2611 AAGATGAGCCACCAAGACCAAGTGAACAAATTCGCTGAGAACTGAGAGCAAGATTG 2670
Qy 2701 CGGAGAGTCACTGAGACAGAGAGGAGAACTGAGCTCAAGGCGCAGCTCAAGAG 2760
Db 2671 CGGAGAGTCACTGAGACAGAGAGGAGAACTGAGCTCAAGGCGCAGCTCAAGAG 2730
Qy 2761 CTACAGCTCTCCCTGAGAGGCGGAGTCAAGTTGAACAGCCCTGACAGCTCAAGGCG 2820
Db 2731 CTACAGCTCTCCCTGAGAGGCGGAGTCAAGTTGAACAGCCCTGACAGCTCAAGGCG 2790
Qy 2821 GCCCTGAGAGGCGGAGTTCGAGAGGCGGAGAGCTGAGAGAGCAACAGCAAGAT 2880
Db 2791 GCCCTGAGAGGCGGAGTTCGAGAGGCGGAGAGCTGAGAGAGCAACAGCAAGAT 2850
Qy 2881 GAAAGAGAGATCAGGCACTCAGGCACTAGAGATGAATCAAGCGCAAAATTTGATCT 2940
Db 2851 GAAAGAGAGATCAGGCACTCAGGCACTAGAGATGAATCAAGCGCAAAATTTGATCT 2910
Qy 2941 CTTCGTAACGCTGTAATCTGTAATCAAGCTGAGAGAGCTTAAACAGCTGACCGAG 3000
Db 2911 CTTCGTAACGCTGTAATCTGTAATCAAGCTGAGAGAGCTTAAACAGCTGACCGAG 2970
Qy 3001 GACAAAGCTGAGACTCAACACCAAAATCTTCACTTCAAAACACTGATGAGGCTCT 3060
Db 2971 GACAAAGCTGAGACTCAACACCAAAATCTTCACTTCAAAACACTGATGAGGCTCT 3030
Qy 3061 GGGGCCAAGCAGAGATTTGTAACACTGGAAGTGAATGACCATCTCCCGGAGATC 3120
Db 3031 GGGGCCAAGCAGAGATTTGTAACACTGGAAGTGAATGACCATCTCCCGGAGATC 3090
Qy 3121 ACGGAAACGAGATGAGCTTACAGCCAGAGAAAGAAAGATGAGGCTTGAAGACCAAG 3180
Db 3091 ACGGAAACGAGATGAGCTTACAGCCAGAGAAAGAAAGATGAGGCTTGAAGACCAAG 3150
Qy 3181 TGCACCATCTGAGAGAAACAGTCAATTTGAGAGCCCTTAAACGATGAGCTGTAGAA 3240
Db 3151 TGCACCATCTGAGAGAAACAGTCAATTTGAGAGCCCTTAAACGATGAGCTGTAGAA 3210
Qy 3241 AAAGAGCGGAGTGGAGGCGCTGAGAGGCGGCTGAGAGTGAAGAAATCCAGATTAG 3300
Db 3211 AAAGAGCGGAGTGGAGGCGCTGAGAGGCGGCTGAGAGTGAAGAAATCCAGATTAG 3270

QY 3301 TGTGGGTTGAGAGCTGCGAGAGATGCTGGAACCCAGAAACAGAGCAGGGCGAGAGCC 3360
DB 3271 TGTGGGTTGAGAGCTGCGAGAGATGCTGGAACCCAGAAACAGAGCAGGGCGAGAGCC 3330
QY 3361 GATCAGCGGATCAACCGAGTCTCGCAGAGTGTGAGAGTGGCAGTGAAGAGCAAGAGCT 3420
DB 3331 GATCAGCGGATCAACCGAGTCTCGCAGAGTGTGAGAGTGGCAGTGAAGAGCAAGAGCT 3390
QY 3421 GAGATTCCTGCTGCGAGAGAGCTCTTCAAAGAGCAAGAGTGAAGGCGGAGAGCTCTCT 3480
DB 3391 GAGATTCCTGCTGCGAGAGAGCTCTCAAAGAGCAAGAGTGAAGGCGGAGAGCTCTCT 3450
QY 3481 GACAGCTCATGATGCTGGAAGAGATGCTATGCTTGAATGATGCTGCGAGAGCTTA 3540
DB 3451 GACAGCTCATGATGCTGGAAGAGATGCTATGCTTGAATGATGCTGCGAGAGCTTA 3510
QY 3541 CAGCAGAGCTGAGAGCTGGAACGAGAGCTCAAAGAGGCTTCTGGAAGAGCAAGCCAA 3600
DB 3511 CAGCAGAGCTGAGAGCTGGAACGAGAGCTCAAAGAGGCTTCTGGAAGAGCAAGCCAA 3570
QY 3601 TTACAGAGCAGAGATGAGACTGCGAAGAAATCACTTTCCGCTGATCTAAGAGCTGCA 3660
DB 3571 TTACAGAGCAGAGATGAGACTGCGAAGAAATCACTTTCCGCTGATCTAAGAGCTGCA 3630
QY 3661 GAAAGCTCTAGATCGGGCTGATCTACTGAGAGCAGAAAGAGTGAAGTCTGAGATCAAGCTG 3720
DB 3631 GAAAGCTCTAGATCGGGCTGATCTACTGAGAGCAGAAAGAGTGAAGTCTGAGATCAAGCTG 3690
QY 3721 GAAACATTCAGGTTCTCTATTCATGAGAAAGGTGAAGTGAAGGACTATTTCTCAA 3780
DB 3691 GAAACATTCAGGTTCTCTATTCATGAGAAAGGTGAAGTGAAGGACTATTTCTCAA 3750
QY 3781 CAAACCAACCTGATGATTTCTGCAAGCCAAATGAGACCACTGCTAAAGAAAGAAA-- 3838
DB 3751 CAAACCAACCTGATGATTTCTGCAAGCCAAATGAGACCACTGCTAAAGAAAGAAAAG 3810
QY 3839 -----AGTTCTCTGAGTAC 3855
DB 3811 GGTATTTAGTCAGCGAAGAAAGAGACCTGCTTACCAACAGGTTCTCTGAGTAC 3870
QY 3856 AATGAGCTGAGAGCTGGCCCTGAGAGAGAGAAAGCTGCTGAGAGCTGAGAGAGAGCC 3915
DB 3871 AATGAGCTGAGAGCTGGCCCTGAGAGAGAGAAAGCTGCTGAGAGCTGAGAGAGAGCC 3930
QY 3916 CTTGAGAGAGCCCGCATGAGCTCGGTCGCGCCGAGAGAGAGAGTCCCAACCGCAAGCA 3975
DB 3931 CTTGAGAGAGCCCGCATGAGCTCGGTCGCGCCGAGAGAGAGTCCCAACCGCAAGCA 3990
QY 3976 ACGAGCAACCCAGCCATCCAGAGCCAGCCAGCCAGAGAGAGAGATCCGCAATGTCGCC 4035
DB 3991 ACGAGCAACCCAGCCATCCAGAGCCAGCCAGCCAGAGAGAGATCCGCAATGTCGCC 4050
QY 4036 ATCTGCGGTGCGCAGAGAGCAGAGCCAGTGCATGAGCTGCTGCGCCGATCCAGC 4095
DB 4051 ATCTGCGGTGCGCAGAGAGCAGAGCCAGTGCATGAGCTGCTGCGCCGATCCAGC 4110
QY 4096 CGCAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTTAAGAAACGATGAC 4155
DB 4111 CGCAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTTAAGAAACGATGAC 4170
QY 4156 CACAATATTCCTCAACCGATTCAGAGTGAAGTGAACATGCGAGCAAAAGTGTGCTG 4215
DB 4171 CACAATATTCCTCAACCGATTCAGAGTGAAGTGAACATGCGAGCAAAAGTGTGCTG 4230
QY 4216 TGTCTGATACCGTGAATTTGAGAGCCAGGATCCAAATGCTCGAATGTCAAGTGATG 4275
DB 4231 TGTCTGATACCGTGAATTTGAGAGCCAGGATCCAAATGCTCGAATGTCAAGTGATG 4290
QY 4276 TGTCAACCCCAAGTGTCAAGTGTGCGAGCCAGCTGCGGCTGCTGCGAATATAGCC 4335
DB 4291 TGTCAACCCCAAGTGTCAAGTGTGCGAGCCAGCTGCGGCTGCTGCGAATATAGCC 4350

QY 4336 AACACCTTCAACCGAGGCTTCTGCGGTGAGCAAAATGAACTCCCAAGTCTCAGACCAAG 4395
DB 4351 AACACCTTCAACCGAGGCTTCTGCGGTGAGCAAAATGAACTCCCAAGTCTCAGACCAAG 4410
QY 4396 GAGCCAGAGAGCTTCACTGAGAGAGTGAAGTGAAGTGGCCAGAGAAATPACAAAGCA 4455
DB 4411 GAGCCAGAGAGCTTCACTGAGAGAGTGAAGTGAAGTGGCCAGAGAAATPACAAAGCA 4470
QY 4456 GAGCAGCAAGGCTGGAGACAGAAATGATGCTGAGAGGATCAAAAGTCTCTATTTAT 4515
DB 4471 GAGCAGCAAGGCTGGAGACAGAAATGATGCTGAGAGGATCAAAAGTCTCTATTTAT 4530
QY 4516 GACAAATGAAGCCAGAGAGCTGGAACAGAGGCGGAGTGAAGAAATTTAGCTGTGCTCCC 4575
DB 4531 GACAAATGAAGCCAGAGAGCTGGAACAGAGGCGGAGTGAAGAAATTTAGCTGTGCTCCC 4590
QY 4576 GACGAGAGTATCTATTCATGAGTCCGTTGCTGCTGCACTGCAAAATPACAGCAAA 4635
DB 4591 GACGAGAGTATCTATTCATGAGTCCGTTGCTGCTGCAAACTGCAAAATPACAGCAAA 4650
QY 4636 GAGAGTGTCCATPACATPACTGAGATGCAATCTCACCCGACACCACTGCTGCGCCGAG 4695
DB 4651 GCA----- 4653
QY 4696 AGAACCTTCTACTTCTAGCTTCCAGCTTCCCTGACAAACAGCGCTGGGTCAACGCTTA 4755
DB 4654 ----- 4653
QY 4756 GAATCAGTTGTGCGAGGTGGAGAGTTCTAGGAGAAAGCAAGAGCTGATGCTTAACTG 4815
DB 4654 -----GAAAGAGCAAGAGCTGATGCTTAACTG 4680
QY 4816 CTTGAGAACTCCCTCTGTAACCTGAGAGATGATGACCTGATGACATGACGACGCTG 4875
DB 4681 CTTGAGAACTCCCTCTGTAACCTGAGAGATGATGACCTGATGACATGACGACGCTG 4740
QY 4876 CCGTTGAGTACCAAGTGTGTTGTTGGGACCGAGAGAGGCTCTACGCTGATATGTC 4935
DB 4741 CCGTTGAGTACCAAGTGTGTTGTTGGGACCGAGAGAGGCTCTACGCTGATATGTC 4800
QY 4936 TTGAGAAACTCCCTTAACCCATGTCGCAAGAAATTTGAGAGAGTTCCTCAAAATTTATATC 4995
DB 4801 TTGAGAAACTCCCTTAACCCATGTCGCAAGAAATTTGAGAGAGTTCCTCAAAATTTATATC 4860
QY 4996 AAGAGCTTGAAGAGCTACTATGATAGAGAGAGAGCGGACCTGTGTGTTGAGC 5055
DB 4861 AAGAGCTTGAAGAGCTACTATGATAGAGAGAGAGCGGACCTGTGTGTTGAGC 4920
QY 5056 GTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCCAGCCGACATCTCA 5115
DB 4921 GTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCCAGCCGACATCTCA 4980
QY 5116 CCCAACATTTTGAAGAGTGTCAAGAGGCTGCCACTGTTTGGGGCAGGCAAGTTGAGAC 5175
DB 4981 CCCAACATTTTGAAGAGTGTCAAGAGGCTGCCACTGTTTGGGGCAGGCAAGTTGAGAC 5040
QY 5176 GGGCTGTGATCTGTGACAGCCATGCCAGCAAAATGCTCATTTCTCCGCTCAACAGAAAC 5235
DB 5041 GGGCTGTGATCTGTGACAGCCATGCCAGCAAAATGCTCATTTCTCCGCTCAACAGAAAC 5100
QY 5236 CTCAGCAAAATPCTGATCCGAGAAAGATAGAGACTCAGAGCCCTGACGCTGATCCAC 5295
DB 5101 CTCAGCAAAATPCTGATCCGAGAAAGATAGAGACTCAGAGCCCTGACGCTGATCCAC 5160
QY 5296 TTACCAATTTACAGTATCTCTATTGGAACCAATTAATTTACGAATGACATGAAGAG 5355
DB 5161 TTACCAATTTACAGTATCTCTATTGGAACCAATTAATTTACGAATGACATGAAGAG 5220
QY 5356 TACAGCTCGAGAGATTCCTGGAATAAGATGACATTCCTTGGGACCTGCTGTGTTGCC 5415
DB 5221 TACAGCTCGAGAGATTCCTGGAATAAGATGACATTCCTTGGGACCTGCTGTGTTGCC 5280
QY 5416 GCGCTCTTCCAGAGCTTCCCTGTCTCAATCTGTGAGAGTGAACAGGACGAGGACGAGAG 5475

Db 5281 GCTCTTCCAA CAGCTTCCCTGCTCTCAATCCGTGACGGTGAACACGACGGGACACGAGAG 5340
 QY 5476 GAGTACTTGCTGTTTCCAGAAATTGAGAGTTTCGGAATCTTTCGGAAGACGTAGC 5535
 Db 5341 GAGTACTTGCTGTTTCCAGAAATTGAGAGTTTCGGAATCTTTCGGAAGACGTAGC 5400
 QY 5536 CGCACAGACGATCTCAATGAGAGTGCCTTACCTTTGAGCTTTGAGCTTACAGAGAACCTTAT 5595
 Db 5401 CGCACAGACGATCTCAATGAGAGTGCCTTACCTTTGAGCTTTGAGCTTACAGAGAACCTTAT 5460
 QY 5596 CTGTTTGTGACCCCACTTCACTCACTGAGAGTAATGAGATCCGAGACGCTCTCTACGCA 5655
 Db 5461 CTGTTTGTGACCCCACTTCACTCACTGAGAGTAATGAGATCCGAGACGCTCTCTACGCA 5520
 QY 5656 GGGACCCCTGCGCCGAGCTACCTGAGACATCCCGAACCCGCGCTACTGAGGCGCTGACATT 5715
 Db 5521 GGGACCCCTGCGCCGAGCTACCTGAGACATCCCGAACCCGCGCTACTGAGGCGCTGACATT 5680
 QY 5716 TCCTCAGAGAGGATTTACTTGGCGTCTCATACCAAGATTAATTAGGGTCAATTGCTGC 5775
 Db 5581 TCCTCAGAGAGGATTTACTTGGCGTCTCATACCAAGATTAATTAGGGTCAATTGCTGC 5640
 QY 5776 AAGGGAAACCTCGTGAAGAGATCCGGGACCTGAGACCAACGCGGGGCGCTCCACCTCCGCG 5835
 Db 5641 AAGGGAAACCTCGTGAAGAGATCCGGGACCTGAGACCAACGCGGGGCGCTCCACCTCCGCG 5700
 QY 5836 AGCAGCCCAACAAGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 5895
 Db 5701 AGCAGCCCAACAAGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 5760
 QY 5896 TCACGCGGAGCGGCGGCGGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 5955
 Db 5761 TCACGCGGAGCGGCGGCGGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 5820
 QY 5956 TACCGCGGAGCGGCGGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 6015
 Db 5821 TACCGCGGAGCGGCGGAGAGGCCCAACGATGACAGAGACGATCACCACAGCGGCTGAGC 5880
 QY 6016 GAGAGTCTCCCGGCGGATCTCAAGACCGGAGAGAGCGGCTGAGAGAGCGGCTGAGAG 6075
 Db 5881 GAGAGTCTCCCGGCGGATCTCAAGACCGGAGAGAGCGGCTGAGAGAGCGGCTGAGAG 5940
 QY 6076 GAGAGTCTCCCGGCGGATCTCAAGACCGGAGAGAGCGGCTGAGAGAGCGGCTGAGAG 6135
 Db 5941 GAGAGTCTCCCGGCGGATCTCAAGACCGGAGAGAGCGGCTGAGAGAGCGGCTGAGAG 6000
 QY 6136 AACCAAGTCTGAGACCAAGTCTTTC 6188
 Db 6001 AACCAAGTCTGAGACCAAGTCTTTC 6023
 Db 6001 AACCAAGTCTGAGACCAAGTCTTTC 6023
 RESULT 11
 AAC77568
 ID AAC77568 standard; cDNA; 6609 BP.
 XX AAC77568;
 AC AAC77568;
 XX AAC77568;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF3123 polynucleotide sequence SEQ ID NO:6245.
 XX
 XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnerability; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antineoplastic; antidiabetic;
 KW antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive; ss.
 OS Homo sapiens.
 XX MO20058473-A2.
 PN 05-OCT-2000.
 PD 31-MAR-2000; 2000MO-US08621.
 PF 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach M,
 PI WPI; 2000-602362/57.
 DR P-PSDB; AAB43359.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 5; Page 5429-5433; 550pp; English.
 CC AAC7446 to AAC77606 encode the proteins given in AAB4023 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerability;
 CC antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;
 CC antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, immunodeficiency (SCID), AIDS, viral,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 6609 BP; 1694 A; 1779 C; 1646 G; 1490 T; 0 other;
 Query Match 62.5%; Score 3855; DB 21; Length 6609;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 3858; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2303 AAGTGTGACCAATCAGATTAAGAAAGACCTGCTGACAGAGACATCTGGAACATCA 2362
 Db 299 AGGTGTGACCAATCAGATTAAGAAAGACCTGCTGACAGAGACATCTGGAACATCA 358
 QY 2363 TCAGAGACAG 2422
 Db 359 TCAGAGACAG 418
 QY 2423 TATCATCTCTATGATTCAGATTCAGATTCCTGGAACAGAGATTGGAACATCTG 2482
 Db 419 TATCATCTCTATGATTCAGATTCAGATTCCTGGAACAGAGATTGGAACATCTG 478
 QY 2483 AAGCCATTAACCTGAGCAATATAGAGCTTTTACCAAGAAAGAACTGAAGGCCCAAG 2542
 Db 479 AAGCCATTAACCTGAGCAATATAGAGCTTTTACCAAGAAAGAACTGAAGGCCCAAG 538

QY 2543 AAGAGATGATTTCTGAATCTGAGGCAAGAAAATTTTACTCTGGAGACACAGGCTGGAACT 2602
DB 539 AAGAGATGATTTCTGAATCTGAGGCAAGAAAATTTTACTCTGGAGACACAGGCTGGAACT 598
QY 2603 TGGAGGCCCAAGAACCGAAAATCTGGAGAGCAGCTGGAGAAATGAGCCACCAAGACACA 2662
DB 599 TGGAGGCCCAAGAACCGAAAATCTGGAGAGCAGCTGGAGAAATGAGCCACCAAGACACA 658
QY 2663 GTGACAAAGATCCGCTGCTGGAACTGGAGCAAGATTTCCGAGAGTCACTAGAGCAGC 2722
DB 659 GTGACAAAGATCCGCTGCTGGAACTGGAGCAAGATTTCCGAGAGTCACTAGAGCAGC 718
QY 2723 AGGAGAGAGAACTGGAGCTTCAAGGCGCAGCTCACAGCTTACAGCTCTCCCTGAGAGC 2782
DB 719 AGGAGAGAGAACTGGAGCTTCAAGGCGCAGCTCACAGCTTACAGCTCTCCCTGAGAGC 778
QY 2783 GCGAGTCACAGTTGACAGCCCTGACAGGCTGCACGCGGCGCCCTGGAGAGCAGCTTCCG 2842
DB 779 GCGAGTCACAGTTGACAGCCCTGACAGGCTGCACGCGGCGCCCTGGAGAGCAGCTTCCG 838
QY 2843 AGGCGAAGACAGAGCTGGAGAGAGACACAGCAGAGCTGAGAGAGATCCAGGCACTCA 2902
DB 839 AGGCGAAGACAGAGCTGGAGAGAGACACAGCAGAGCTGAGAGAGATCCAGGCACTCA 898
QY 2903 CGGACATAGAGATGAAATCCAGCGCAAAATTTGATGCTCTTCCGTAAACAGCTTACTGTAA 2962
DB 899 CGGACATAGAGATGAAATCCAGCGCAAAATTTGATGCTCTTCCGTAAACAGCTTACTGTAA 958
QY 2963 TCACAGACCTGGAGAGCAGCTTAAACAGCTGACCGAGACCAACGCTGAACTCAACACC 3022
DB 959 TCACAGACCTGGAGAGCAGCTTAAACAGCTGACCGAGAGCAACGCTGAACTCAACACC 1018
QY 3023 AAAACCTTACTTGTTCACAAACAACTGATGAGGCTTCTGGCCGCAACGACGAGATTGTAC 3082
DB 1019 AAAACCTTACTTGTTCACAAACAACTGATGAGGCTTCTGGCCGCAACGACGAGATTGTAC 1078
QY 3083 AACTGGAGATGAAAGTGAACCACTCTCCGCGGAGATCAAGGAAACGAGATTCAGCTTAA 3142
DB 1079 AACTGGAGATGAAAGTGAACCACTCTCCGCGGAGATCAAGGAAACGAGATTCAGCTTAA 1138
QY 3143 CCAGCCAGAGCAAAAGATGAGAGGCTCTGAAGACCAAGTGCACATGCTGAGAGAAACAG 3202
DB 1139 CCAGCCAGAGCAAAAGATGAGAGGCTCTGAAGACCAAGTGCACATGCTGAGAGAAACAG 1198
QY 3203 TCATGGAATTTGGAGGCGCTTAAACGATGAGCTCTAGAAAAAGAGGCGCAGTGGAGGCT 3262
DB 1199 TCATGGAATTTGGAGGCGCTTAAACGATGAGCTCTAGAAAAAGAGGCGCAGTGGAGGCT 1258
QY 3263 GGAGGAGCGTCTCTGGGTGATGAGAAATCCAGTTTGAAGTGTGGGTTGAGAGCTGCAGA 3322
DB 1259 GGAGGAGCGTCTCTGGGTGATGAGAAATCCAGTTTGAAGTGTGGGTTGAGAGCTGCAGA 1318
QY 3323 GATCTCTGAGACACCGAAGAAACAGACAGAGGCGAGACCGATACCGAGTCTC 3382
DB 1319 GATCTCTGAGACACCGAAGAAACAGACAGAGGCGAGACCGATACCGAGTCTC 1378
QY 3383 GCGAGTGTGGAGTGTGACAGTGAAGAGCAAAAGCTGAGATTTCTCGTCTGACAGCAG 3442
DB 1379 GCGAGTGTGGAGTGTGACAGTGAAGAGCAAAAGCTGAGATTTCTCGTCTGACAGCAG 1438
QY 3443 CTCTCAAGAGCAGAAAGCTGAAGGCGGAGGCTCTGACCAAGCTCAATGACCTGAGAG 3502
DB 1439 CTCTCAAGAGCAGAAAGCTGAAGGCGGAGGCTCTGACCAAGCTCAATGACCTGAGAG 1498
QY 3503 AGAAGCATGCTATGTTGAATGAATGCCGAAAGCTTACAGCAGAAAGCTGAGAGCTGAGAC 3562
DB 1499 AGAAGCATGCTATGTTGAATGAATGCCGAAAGCTTACAGCAGAAAGCTGAGAGCTGAGAC 1558
QY 3563 GAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCCAAATTACAGCAGAGATGAGACCTGC 3622
DB 1559 GAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCCAAATTACAGCAGAGATGAGACCTGC 1618

QY 3623 AGAAAAATCAATTTTCCGTCTGACTCAGAGACTGCAAGAAAGCTTAGATCGGCTGATC 3682
DB 1619 AGAAAAATCAATTTTCCGTCTGACTCAGAGACTGCAAGAAAGCTTAGATCGGCTGATC 1678
QY 3683 TACTGAAAGACAGAAAGAGTGAAGTCTGAGTATCAGCTGGAACCAATTCAGTTCTTAAT 3742
DB 1679 TACTGAAAGACAGAAAGAGTGAAGTGAATCAGCTGGAACCAATTCAGTTCTTAAT 1738
QY 3743 CTGATGAAAGAGTGAAGAAATGGAAGGCACTATTTCACAAACCAATCTCATTAATTTTC 3802
DB 1739 CTGATGAAAGAGTGAAGAAATGGAAGGCACTATTTCACAAACCAATCTCATTAATTTTC 1798
QY 3803 TGCAGCCAAATATGACCAACTCTGCTTAAAGAAAAAGTTTCCCTCAGATGACATGAGC 3862
DB 1799 TGCAGCCAAATATGACCAACTCTGCTTAAAGAAAAAGTTTCCCTCAGATGACATGAGC 1858
QY 3863 TGAAGCTGCGCTTGGAGAGAGAGAAAGCTGCTGTGACAGCTAGAGAAAGCCCTTGAGA 3922
DB 1859 TGAAGCTGCGCTTGGAGAGAGAGAAAGCTGCTGTGACAGCTAGAGAAAGCCCTTGAGA 1918
QY 3923 AGACCCGCAATCGAGCTCCGCTCCGCGGAGAGAAAGTGCACCCGTAAGCAAGCAAC 3982
DB 1919 AGACCCGCAATCGAGCTCCGCTCCGCGGAGAGAAAGTGCACCCGTAAGCAAGCAAC 1978
QY 3983 ACCCAGACCATCAGAGCCAGCCAGCCAGAGCAGAGATGCGCATGTCGCAATCGTGC 4042
DB 1979 ACCCAGACCATCAGAGCCAGCCAGCCAGAGCAGAGATGCGCATGTCGCAATCGTGC 2038
QY 4043 GGTCCGCAAGACACAGCCCAAGTGCATGAGCTGCTGGCCCGCCATCCAGCCGAGAA 4102
DB 2039 GGTCCGCAAGACACAGCCCAAGTGCATGAGCTGCTGGCCCGCCATCCAGCCGAGAA 2098
QY 4103 AGGAGCTTCAACTCCAGAGAAATTAAGTCGGGCTTTAAGGAACCATGACACACATA 4162
DB 2099 AGGAGCTTCAACTCCAGAGAAATTAAGTCGGGCTTTAAGGAACCATGACACACATA 2158
QY 4163 TTCTTCACCGATTCAACGTAGAGACTGAACATGACAGCCCAAAAGTGTGTGTGTGG 4222
DB 2159 TTCTTCACCGATTCAACGTAGAGACTGAACATGACAGCCCAAAAGTGTGTGTGTGG 2218
QY 4223 ATACCGTGACCTTTGAGAGCCAGGCAATCCAAATGTCGAAATGTCAGTGTGTGTACC 4282
DB 2219 ATACCGTGACCTTTGAGAGCCAGGCAATCCAAATGTCGAAATGTCAGTGTGTGTACC 2278
QY 4283 CCAAGTGTCCACGCTTGGCAGCACTTGCGGCTTGGCTGTAATGTCACACACT 4342
DB 2279 CCAAGTGTCCACGCTTGGCAGCACTTGCGGCTTGGCTGTAATGTCACACACT 2338
QY 4343 TCACGAGGCTTGTGCGGTGACAAATGAATCCCAAGTCTCCAGACCAAGAGCCCA 4402
DB 2339 TCACGAGGCTTGTGCGGTGACAAATGAATCCCAAGTCTCCAGACCAAGAGCCCA 2398
QY 4403 GCAGAGCTTCACCTGGAAGGCTGATGAAGGTCGCCAGGAATTAACAGAGACAGC 4462
DB 2399 GCAGAGCTTCACCTGGAAGGCTGATGAAGGTCGCCAGGAATTAACAGAGACAGC 2458
QY 4463 AAGCTGGAAGAGATGACATTTGCTGAGAGGATCAAAAGTCCCTATTAATGACATG 4522
DB 2459 AAGCTGGAAGAGATGACATTTGCTGAGAGGATCAAAAGTCCCTATTAATGACATG 2518
QY 4523 AAGCCAGAAAGCTGAGACAGGCGCGTGAAGAAATTTGAAGCTGCTTCCGACGAGG 4582
DB 2519 AAGCCAGAAAGCTGAGACAGGCGCGTGAAGAAATTTGAAGCTGCTTCCGACGAGG 2578
QY 4583 ATGTATCATTCATGAGGCGCTTGTGCTTCCGAACTCCGAAATTCAGCCAAAGAGATG 4642
DB 2579 ATGTATCATTCATGAGGCGCTTGTGCTTCCGAACTCCGAAATTCAGCCAAAGAGATG 2638
QY 4643 TCCCATCATTCATGAGATGGAATCTCACCCGACACCACTGCTGGCCGGAGAACCC 4702
DB 2639 TCCCATCATTCATGAGATGGAATCTCACCCGACACCACTGCTGGCCGGAGAACCC 2698
QY 4703 TCTACTGTGATGCTCCAGCTTCCCTGACAAACAGCCCTGGGTCAACGCTTTAGATCAG 4762

```

Db      |||||
2699  TCCTACTGCTAGCTCCAGCTCTCCCTGACAAACAGCGCTGGGTACCGCCTTGATGATCAG 2758
Qy      |||||
4763  TTGTCCGAGGTGGGAGAGTTTCTTAGGAAAAAGAGAAAGTGAATGCTTAACTCTTGGA 4822
Db      |||||
2759  TTGTCCGAGGTGGGAGAGTTTCTTAGGAAAAAGAGAAAGTGAATGCTTAACTCTTGGA 2818
Qy      |||||
4823  ACTCCCTGCTGAAACTGGAAGGTGATACCGTCTTAGACATGAACTGACCGCTCCCTTCA 4882
Db      |||||
2819  ACTCCCTGCTGAAACTGGAAGGTGATACCGTCTTAGACATGAACTGACCGCTCCCTTCA 2878
Qy      |||||
4883  GTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4942
Db      |||||
2879  GTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2938
Qy      |||||
4943  ACTCCCTGCTGAAACTGGAAGGTGATACCGTCTTAGACATGAACTGACCGCTCCCTTCA 5002
Db      |||||
2939  ACTCCCTGCTGAAACTGGAAGGTGATACCGTCTTAGACATGAACTGACCGCTCCCTTCA 2998
Qy      |||||
5003  TGGAGAAAGCTAATCATGATATAGCAGAGAAAGCGGGCACTGTGTCTTGTGTGAGTGAAG 5062
Db      |||||
2999  TGGAGAAAGCTAATCATGATATAGCAGAGAAAGCGGGCACTGTGTCTTGTGTGAGTGAAG 3058
Qy      |||||
5063  AAGTGAACAGTCCCTGGGAGGAGTCCGACCTGCTGCGAGCCCGAGCATCTCACCGAACA 5122
Db      |||||
3059  AAGTGAACAGTCCCTGGGAGGAGTCCGACCTGCTGCGAGCCCGAGCATCTCACCGAACA 3118
Qy      |||||
5123  TTTTGAAGCTGTCAAGGAGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5182
Db      |||||
3119  TTTTGAAGCTGTCAAGGAGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3178
Qy      |||||
5183  GCATCTGTGACGCTATCCCAAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5242
Db      |||||
3179  GCATCTGTGACGCTATCCCAAGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3238
Qy      |||||
5243  AATAGTGCATCCGGAAGAGATGAGACCTCAGAGCCCTGACCTGATCACTTGCACCA 5302
Db      |||||
3239  AATAGTGCATCCGGAAGAGATGAGACCTCAGAGCCCTGACCTGATCACTTGCACCA 3298
Qy      |||||
5303  ATTAGATATCTCTCATTTGGAACCAATTAATTTTACGAAATGCAATGAGACAGTACACGC 5362
Db      |||||
3299  ATTAGATATCTCTCATTTGGAACCAATTAATTTTACGAAATGCAATGAGACAGTACACGC 3358
Qy      |||||
5363  TCGAGGAATCTCTGATTAAGATGACATCTCTGAGACCTGCTGTGTGTGTGTGTGTGTGTGT 5422
Db      |||||
3359  TCGAGGAATCTCTGATTAAGATGACATCTCTGAGACCTGCTGTGTGTGTGTGTGTGTGTGT 3418
Qy      |||||
5423  CCAACAGCTTCCCTGTCTCAATGCTGAGGTGAACAGCGAGGAGCAAGCAAGAGTACT 5482
Db      |||||
3419  CCAACAGCTTCCCTGTCTCAATGCTGAGGTGAACAGCGAGGAGCAAGCAAGAGTACT 3478
Qy      |||||
5483  TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5542
Db      |||||
3479  TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3538
Qy      |||||
5543  ACGATCTCAAGTGAAGTGGCTTACCTTTGCTTGTGCTTACAGAGAACCTTATCTGTTG 5602
Db      |||||
3539  ACGATCTCAAGTGAAGTGGCTTACCTTTGCTTGTGCTTACAGAGAACCTTATCTGTTG 3598
Qy      |||||
5603  TGAACCACTTCACTCACTGAAAGTAAATGAGATCCAGGCAAGCTCTCTACAGAGAGCC 5662
Db      |||||
3599  TGAACCACTTCACTCACTGAAAGTAAATGAGATCCAGGCAAGCTCTCTACAGAGAGCC 3658
Qy      |||||
5663  CTGCGGAGGAGTACTGAGCATCCCGAACCCGAGCTACCTGGGAGCTGCTGCTGCTGCTGCT 5722
Db      |||||
3659  CTGCGGAGGAGTACTGAGCATCCCGAACCCGAGCTACCTGGGAGCTGCTGCTGCTGCTGCT 3718
Qy      |||||
5723  GAGGATTTAATCTTGGGCTCTCTCATACCAAGATTAATTAAGGATCATTTGTGTGAGGAA 5782
Db      |||||
3719  GAGGATTTAATCTTGGGCTCTCTCATACCAAGATTAATTAAGGATCATTTGTGTGAGGAA 3778
Qy      |||||
5783  ACCCTGTGAAGAGTCCGAGCTGAACCAACCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5842

```

```

Db      |||||
3779  ACCCTGTGAAGAGTCCGAGCACTGAACACCAACCGGGGCTGCTGCTGCTGCTGCTGCTGCT 3838
Qy      |||||
5843  CCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5902
Db      |||||
3839  CCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3898
Qy      |||||
5903  CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5962
Db      |||||
3899  CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3958
Qy      |||||
5963  AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6022
Db      |||||
3959  AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4018
Qy      |||||
6023  CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6082
Db      |||||
4019  CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4078
Qy      |||||
6083  GCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6142
Db      |||||
4079  GCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4138
Qy      |||||
6143  TCTGGAGACCACTTCTCACTATTA 6165
Db      |||||
4139  TCTGGAGACCACTTCTCACTATTA 4161

RESULT 12
ABR08361
ID  ABR08361 standard; cDNA, 3131 BP.
AC  ABA08361;
DT  11-JAN-2002 (first entry)
DE  Human RHO/RAC effector homologue-encoding cDNA. SEQ ID NO:137.
XX
XX
XX  Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX  haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX  inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX  proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX  myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX  chronic inflammatory condition; proliferative retinopathy;
XX  atherosclerosis; coronary heart disease; arterial ischaemia;
XX  bone disorder; osteoporosis; vascular growth disorder;
XX  tissue regeneration; wound healing; infection; immune disorder;
XX  cell culture; drug screening; gene therapy; antiinflammatory;
XX  antiaslarmic; antiarthritic; haemostatic; antiarteriosclerotic;
XX  cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX  antifungal; vulnery; antulcer; ss.
XX
XX  Homo sapiens.
XX
XX  WO200157188-A2.
XX
XX  09-AUG-2001.
XX
XX  05-FEB-2001; 2001WO-US03800.
XX
XX  03-FEB-2000; 2000US-0496914.
XX  27-APR-2000; 2000US-0560875.
XX
XX  (HYSB-) HYSBQ INC.
XX
XX  Tang YT, Liu C, Dymnac RT;
XX
XX  WPI; 2001-457740/49.
XX  P-PDB; ABB11117.
XX
XX  Human proteins and DNA encoding sequences useful for preventing,
XX  treating or ameliorating a medical condition in a mammalian subject
XX  e.g. arthritis and cancer -
XX

```

PS Claim 1; Page 387-388; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities, stem cell growth factor activity;
CC hematopoietic regulatory activity, tissue growth activity;
CC immunomodulatory activity, activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.
XX
XX
SQ Sequence 3131 BP; 861 A; 835 C; 806 G; 629 T; 0 other;

Query Match 47.5%; Score 2928.2; DB 22; Length 3131;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2905 GCACATGAGATGAAATCCAGCGCAATTGATGCTCTTCTGTAACAGCTGTAAGTATC 2964
DB 1 GCACATGAGATGAAATCCAGCGCAATTGATGCTCTTCTGTAACAGCTGTAAGTATC 60
QY 2965 ACAGACCTGGAGAGAGAGCTTAAACAGCTGACCGAGAGACAGCTTAACTCAACACCA 3024
DB 61 ACAGACCTGGAGAGAGAGCTTAAACAGCTGACCGAGAGACAGCTTAACTCAACACCA 120
QY 3025 AACTCTACTCTTCCAAACAACTCGATGAGGCTTTCGGCGCAACGAGAGATTGACAA 3084
DB 121 AACTCTACTCTTCCAAACAACTCGATGAGGCTTTCGGCGCAACGAGAGATTGACAA 180
QY 3085 CTCGGAAGTGAAGTGAACATCTTCGCCCGGAGATACGGAACGAGAGATGACGTTACC 3144
DB 181 CTCGGAAGTGAAGTGAACATCTTCGCCCGGAGATACGGAACGAGAGATGACGTTACC 240
QY 3145 AGCCAGAGCAAGAGATGAGGCTCTGAGAGACAGCTGACCACTGCTGAGAGAACAGGTC 3204
DB 241 AGCCAGAGCAAGAGATGAGGCTCTGAGAGACAGCTGACCACTGCTGAGAGAACAGGTC 300
QY 3205 ATGGAATTGAGAGCCCTTAAACGATGAGCTGTAAGAAAAAGAGGCGAGTGGAGGCTTGG 3264
DB 301 ATGGAATTGAGAGCCCTTAAACGATGAGCTGTAAGAAAAAGAGGCGAGTGGAGGCTTGG 360
QY 3265 AGAAGGCTCTGCGGTATGAGAAATCCAGATTGATGTCGCGGTGAGAGAGCTGACAGAG 3324
DB 361 AGAAGGCTCTGCGGTATGAGAAATCCAGATTGATGTCGCGGTGAGAGAGCTGACAGAG 420

QY 3325 ATGCTGACACCCAGAAACAGAGCGAGCGACGCGATGAGCGGATCACCGAGTCTCCG 3384
DB 421 ATGCTGACACCCAGAAACAGAGCGAGCGACGCGATGAGCGGATCACCGAGTCTCCG 480
QY 3385 CAGGTGGTGGAGCTGCGAGTGAAGAGACCAAGAGTGAATTTCTGCTGCGACAGAGCT 3444
DB 481 CAGGTGGTGGAGCTGCGAGTGAAGAGACCAAGAGTGAATTTCTGCTGCGACAGAGCT 540
QY 3445 CTCAAGAGCAAGAGCTGAAGGCCGAGAGGCTCTGACAACTCAATGACTGAGAGAG 3504
DB 541 CTCAAGAGCAAGAGCTGAAGGCCGAGAGGCTCTGACAACTCAATGACTGAGAGAG 600
QY 3505 AAGATGCTATGCTTGAATGAATGCCGAAAGCTTACAGAGAGCTGAGACTGAAGCA 3564
DB 601 AAGATGCTATGCTTGAATGAATGCCGAAAGCTTACAGAGAGCTGAGACTGAAGCA 660
QY 3565 GAGCTCAACAGAGGCTTCTGAAAGCAAGCCAAATTAAGCAGCAGATGACCTGACAG 3624
DB 661 GAGCTCAACAGAGGCTTCTGAAAGCAAGCCAAATTAAGCAGCAGATGACCTGACAG 720
QY 3625 AAAAATCAATTTCCGCTGCTGACTCAAGAGCTGACAGAGCTTAAGTGGGCTGATCTA 3684
DB 721 AAAAATCAATTTCCGCTGCTGACTCAAGAGCTGACAGAGCTTAAGTGGGCTGATCTA 780
QY 3685 CTGAAGACAGAAAGATGAGCTTGAAGTATCAGCTGAGAAACATTGAGTTCTTATCT 3744
DB 781 CTGAAGACAGAAAGATGAGCTTGAAGTATCAGCTGAGAAACATTGAGTTCTTATCT 840
QY 3745 CATGAAAAAGTGAATGAAAGGCACTATTTCTCAACAAACAAACTCATGATTTTCTG 3804
DB 841 CATGAAAAAGTGAATGAAAGGCACTATTTCTCAACAAACAAACTCATGATTTTCTG 900
QY 3805 CAAGCCAAATGAGCAACCTGCTTAAAGAAAAAGTTCTCTGCAAGTCAATGAGACTG 3864
DB 901 CAAGCCAAATGAGCAACCTGCTTAAAGAAAAAGTTCTCTGCAAGTCAATGAGACTG 960
QY 3865 AAGCTGGCCCTGAGAAAGAGAAAGCTCGCTGTGACAGAGTGAAGAGCCCTTCAGAG 3924
DB 961 AAGCTGGCCCTGAGAAAGAGAAAGCTCGCTGTGACAGAGTGAAGAGCCCTTCAGAG 1020
QY 3925 ACCCGCATCGAGCTCCGCTCCGCGGAGAAAGCTGCTCCACCGCAAGCAACGAGACAC 3984
DB 1021 ACCCGCATCGAGCTCCGCTCCGCGGAGAAAGCTGCTCCACCGCAAGCAACGAGACAC 1080
QY 3985 CCAGACCCATCCAGCGCAGCCACCGCAGAGCAGAGATGCTGCTGCTGCTGCTGCTG 4044
DB 1081 CCAGACCCATCCAGCGCAGCCACCGCAGAGCAGAGATGCTGCTGCTGCTGCTGCTG 1140
QY 4045 TGGCCAGAGCACAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4104
DB 1141 TGGCCAGAGCACAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 4105 GAGTCTTCAACTCCAGAGAAATTAAGTCCGAGTCTTAAGAAAGCAATGACCAATAT 4164
DB 1201 GAGTCTTCAACTCCAGAGAAATTAAGTCCGAGTCTTAAGAAAGCAATGACCAATAT 1260
QY 4165 CCTCAGGATTCAGAGTGAAGTGAACATGAGAGCCCAAAAGTGTGTGTGTGAT 4224
DB 1261 CCTCAGGATTCAGAGTGAAGTGAACATGAGAGCCCAAAAGTGTGTGTGTGAT 1320
QY 4225 ACCGTGACCTTGAAGCCCGAGGCACTCAATGCTGCAATGCTGAGTGTGACCC 4284
DB 1321 ACCGTGACCTTGAAGCCCGAGGCACTCAATGCTGCAATGCTGAGTGTGACCC 1380
QY 4285 AAGTGTCTCAAGTGTGCTGAGCCAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 4344
DB 1381 AAGTGTCTCAAGTGTGCTGAGCCAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 1440
QY 4345 ACCGAGGCTTCTGCGCTGAGCAAAATGAATCTCCCAAGTCTCCAGACCAAGAGCCAGC 4404
DB 1441 ACCGAGGCTTCTGCGCTGAGCAAAATGAATCTCCCAAGTCTCCAGACCAAGAGCCAGC 1500

QY	4405	AGACGTTGTCACCTGGAAAGGGTGGATGAAAGTGTCCTCCAGGAATTAACAAACGAGACACGAA	4466
Db	1501	AGCAGGTTCCACCTCGAAGAGGGTGGATGAAAGTGTCCTCCAGGAATTAACAAACGAGACACGAA	1566
QY	4465	GGCTGGACACGGAAGTACATTGGCTGGAGGGATCAAAAGTCCATTTATGACATGAA	4526
Db	1561	GGCTGGACACGGAAGTACATTGGCTGGAGGGATCAAAAGTCCATTTATGACATGAA	1620
QY	4525	GCCAGAGAGCTGGACAGAGGGCGGTGGAAGAAATTTAGAGCTGTGCCTTCCGACGGAGAT	4586
Db	1621	GCCAGAGAGAGCTGGACAGAGGGCGGTGGAAGAAATTTAGAGCTGTGCCTTCCGACGGAGAT	1680
QY	4585	GTATCTATTCAATGTGTCGGTGTGTGCTTCGAACTTCGAAATACGCCAAAGCAGATGTC	4644
Db	1681	GTATCTATTCAATGTGTCGGTGTGTGCTTCGAACTTCGAAATACGCCAAAGCAGATGTC	1740
QY	4645	CCATTCATCTGTAAGATGGAATCTCACTCCGCAACACACTGTGTGGCCCGGAGAACCTTC	4704
Db	1741	CCATTCATCTGTAAGATGGAATCTCACTCCGCAACACACTGTGTGGCCCGGAGAACCTTC	1800
QY	4705	TACTTGTCTAGCTCCGACGTTCCCTGACAAACAGCGCTGGAGTCAACGCTTGAATCAGTT	4764
Db	1801	TACTTGTCTAGCTCCGACGTTCCCTGACAAACAGCGCTGGAGTCAACGCTTGAATCAGTT	1860
QY	4765	GTGCGAGGTGGAGAGTTTCTAGGAGAAAAGCAGAGCTGATGCTAAACTGTCTGGAAC	4824
Db	1861	GTGCGAGGTGGAGAGTTTCTAGGAGAAAAGCAGAGCTGATGCTAAACTGTCTGGAAC	1920
QY	4825	TCCCTGCTGAAATCTGGAAGGTGATGACCGTCTAGACATGAACTGACCGCTCCCTTCAGT	4884
Db	1921	TCCCTGCTGAAATCTGGAAGGTGATGACCGTCTAGACATGAACTGACCGCTCCCTTCAGT	1980
QY	4885	GACCAAGTGTGTGTGTGGTGGGACACGAGAAAGGGCTCTAGCGCCTGAATGTCTGAAAAC	4944
Db	1981	GACCAAGTGTGTGTGTGGTGGGACACGAGAAAGGGCTCTAGCGCCTGAATGTCTGAAAAC	2040
QY	4945	TCCCTAACCCATGTCTCCACGAAATTGGAGCACTTTCCAAATTTATATTATCAAGACCTG	5004
Db	2041	TCCCTAACCCATGTCTCCACGAAATTGGAGCACTTTCCAAATTTATATTATCAAGACCTG	2100
QY	5005	GAGAAAGCTACATCATATAGACAGAGAAAGAGGGGGATCTGTCTTTGTGGACGTGAAGAA	5064
Db	2101	GAGAAAGCTACATCATATAGACAGAGAAAGAGGGGGATCTGTCTTTGTGGACGTGAAGAA	2160
QY	5065	GTGAAACAGTCCCTGGGCCAGTCCCACTGCTGCCACGCCGCACTTCACCCACAAATT	5124
Db	2161	GTGAAACAGTCCCTGGGCCAGTCCCACTGCTGCCACGCCGCACTTCACCCACAAATT	2220
QY	5125	TTTGAAGCTGTCAAGAGGCTGCCTCTGTTTGGGGCAGGACGATTTGAGAACGGGCTCTGC	5184
Db	2221	TTTGAAGCTGTCAAGAGGCTGCCTCTGTTTGGGGCAGGACGATTTGAGAACGGGCTCTGC	2280
QY	5185	ATCTGTGAGACATGACCCACGAAAGTGTCTCATTTCTCGCTTACACGAAAACCTGACAAA	5244
Db	2281	ATCTGTGAGACATGACCCACGAAAGTGTCTCATTTCTCGCTTACACGAAAACCTGACAAA	2340
QY	5245	TACTGCACTCCGAAAAGAGATAGAGACCTCAAGCCCTGACGCTTATTCACCTTCACCAAT	5304
Db	2341	TACTGCACTCCGAAAAGAGATAGAGACCTCAAGCCCTGACGCTTATTCACCTTCACCAAT	2400
QY	5305	TACAGTATCTCATATTGGAAACCAATPAATTTACAGAAATCGACATGAAGCAGTACACGCTC	5364
Db	2401	TACAGTATCTCATATTGGAAACCAATPAATTTACAGAAATCGACATGAAGCAGTACACGCTC	2460
QY	5365	GAGGAATTCCTGGAATGAAGATGACCATTTCTTGGCACCTGTGTGTTTGGCGCTCTTCC	5424
Db	2461	GAGGAATTCCTGGAATGAAGATGACCATTTCTTGGCACCTGTGTGTTTGGCGCTCTTCC	2520
QY	5425	AACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAGCGACAGGGCAGAGGAGAGTACTTGG	5484
Db	2521	AACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAGCGACAGGGCAGAGGAGAGTACTTGG	2580
QY	5485	CTGTGTTTCCACGAATTTGAGTGTGTGTGATTTCTTACGAAAGACGTACGCCACACGAC	5544

Db	2581	CTGTGTTTCCAGGAATTTGGAGTGTCTGTGAATTTCTTAACGGAACAGTACGCCGACAGAC	2640
Qy	5545	GATCTCAAGTGGAGTCGCTTACCTTTGGCCTTTGGCTTACAGAGAACCCCTATCTGTTTGTG	5604
Db	2641	GATCTCAAGTGGAGTCGCTTACCTTTGGCCTTTGGCCTTACAGAGAACCCCTATCTGTTTGTG	2700
Qy	5605	ACCCACTTCACACTCACTGGAAGTAAATGATTCAGAGCAGCCTCTCAAGAGGACCCCT	5664
Db	2701	ACCCACTTCACACTCACTGGAAGTAAATGATTCAGAGCAGCCTCTCTCAAGAGGACCCCT	2760
Qy	5665	GCCCGAGCGTACTTGAATCCCGAACCCGGGCTTACCTCGGGCCCTGCAATTCCTCAAGA	5724
Db	2761	GCCCGAGCGTACTTGAATCCCGAACCCGGGCTTACCTCGGGCCCTGCAATTCCTCAAGA	2820
Qy	5725	GCGAATTAATCTGGCGCTCTCATACCGAGATTAATTAAGGTCATTTGCTGCAAGGAAAC	5784
Db	2821	GCGAATTAATCTGGCGCTCTCATACCGAGATTAATTAAGGTCATTTGCTGCAAGGAAAC	2880
Qy	5785	CTCGTGAAGAGTCCGGACCTGAACACCAACCGGGGCCCTGTCACCTCCCGAG	5837
Db	2881	CTCGTGAAGAGTCCGGACCTGAACACCAACCGGGGCCCTGTCACCTCCCGAG	2933
RESULT 13			
ID	AA155216	standard: DNA; 5261 BP.	
AC	AA155216;		
XX	01-MAY-2003	(first entry)	
DT	Human CRK related DNA sequence, SEQ ID No. 6.		
XX			
XX	Anorectic; hypotensive; cardiact; antilipaeic; cerebroprotective;		
KW	antigout; osteopathic; antiarthritic; cytostatic; antidepressant;		
KW	immunomodulator; antineuric; tranquiliser; antiparkinsonian; nootropic;		
KW	neuroprotective; antiinflammatory; antidiabetic; analgesic;		
KW	human citron rho/rac-interacting kinase; enzyme; CRK; ameliorating;		
KW	obesity; comorbidities; cancer; anorexia; cachexia; bulimia;		
KW	central nervous system disorder; chronic obstructive pulmonary disease;		
KW	diabetes; pain; ds.		
OS	Homo sapiens.		
XX			
XX	WO2003004523-A1.		
XX	16-JAN-2003.		
XX			
XX	28-JUN-2002; 2002WO-EP07156.		
PF			
XX	02-JUL-2001; 2001US-301841P.		
PR	11-DEC-2001; 2001US-338651P.		
PR	25-APR-2002; 2002US-375014P.		
XX			
XX	(FARB) BAYER AG.		
XX			
XX	Zhu Z;		
XX			
XX	WPI; 2003-221576/21.		
XX			
PT	New human citron rho/rac-interacting kinase (CRK) polypeptide and		
PT	polynucleotide, useful in preventing, ameliorating or treating diseases		
PT	associated with human CRK dysfunction, e.g. obesity, diabetes or		
PT	Alzheimer's disease -		
XX			
PS	Disclosure; Fig 6; 237pp; English.		
XX			
CC	The invention relates to an isolated polynucleotide encoding a human		
CC	citron rho/rac-interacting kinase polypeptide. The isolated		
CC	polynucleotide comprises a 6165 or 8603 base pair sequence, given in the		
CC	specification. The human citron rho/rac-interacting kinase (CRK) gene		
CC	polypeptide and polynucleotide are useful in preventing, ameliorating or		

CC treating diseases associated with human CRK dysfunction such as obesity
CC and obesity-associated comorbidities (e.g. hypertension, coronary artery
CC disease, hyperlipidemia, stroke, gout, osteoarthritis, some types of
CC cancer including endometrial, breast, prostate and colon cancer),
CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood
CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's
CC disease), chronic obstructive pulmonary disease, or diabetes. These can
CC also be used to treat pain associated with the disorders. The human CRK
CC polypeptide is also useful in diagnostic assays or in genetic testing.
CC The expression vector or the reagent is useful in preparing a medicament
CC for modulating the activity of a human CRK in a disease, e.g. obesity, a
CC central nervous system disorder, or chronic obstructive pulmonary
CC disease. The fusion protein is useful for generating antibodies against a
CC CRK polypeptide and for use in various assay systems. The methods are
CC useful in producing and detecting the polynucleotide and polypeptide and
CC in screening for agents that modulate the activity of the human CRK
CC polypeptide. This polynucleotide sequence represents a DNA sequence
CC relating to the human CRK protein of the invention.
XX

Sequence 5261 BP; 1309 A; 1465 C; 1283 G; 1204 T; 0 other;

Query Match 45.7%; Score 2818.2; DB 25; Length 5261;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 3343 CAGAGCAGGCGGAGAGCCGATCAGCGGATCAGAGCTTCCAGAGTGTGGAGCTGGCA 3402
Db 1 CAGAGCAGGCGGAGAGCCGATCAGCGGATCAGAGCTTCCAGAGTGTGGAGCTGGCA 60
QY 3403 GTAGAGGAGCAGAGGCTGAGATTCTCGCTCGAGAGGCTCTCAAGAGCGAAAGCTG 3462
Db 61 GTAGAGGAGCAGAGGCTGAGATTCTCGCTCGAGAGGCTCTCAAGAGCGAAAGCTG 120
QY 3463 AAGCGCAGAGCTCTCTGCAAGCTCAATAGCTGAGAGAGAGAGCTATGCTTGA 3522
Db 121 AAGCGCAGAGCTCTCTGCAAGCTCAATAGCTGAGAGAGAGAGCTATGCTTGA 160
QY 3523 ATGAAATGCCGAAAGCTTACGCGAAGCTGGAAGTGAAGCAGAGCTCAAGAGGCTT 3582
Db 181 ATGAAATGCCGAAAGCTTACGCGAAGCTGGAAGTGAAGCAGAGCTCAAGAGGCTT 240
QY 3583 CTGAGAGAGCAGCAGAAATTAACAGCAGAGATGAGACCTGAGAAATCAATTTCCGT 3642
Db 241 CTGAGAGAGCAGCAGAAATTAACAGCAGAGATGAGACCTGAGAAATCAATTTCCGT 300
QY 3643 CTGACTCAAGGACTGCAAGAGCTAGATCGGGCTGATCTACTGGAAGCAGAAAGAGT 3702
Db 301 CTGACTCAAGGACTGCAAGAGCTAGATCGGGCTGATCTACTGGAAGCAGAAAGAGT 360
QY 3703 GACTTGGAGTATAGCTGGAAGAACTTCAAGCTTCTATTTCTATGAAAAGTGAAATG 3762
Db 361 GACTTGGAGTATAGCTGGAAGAACTTCAAGCTTCTATTTCTATGAAAAGTGAAATG 420
QY 3763 GAAGGACTATTTCTCAAGAAACCAATCATGATTTTCTGCAAGCCAAATGAGCA 3822
Db 421 GAAGGACTATTTCTCAAGAAACCAATCATGATTTTCTGCAAGCCAAATGAGCA 480
QY 3823 CTTGCTAAAAAGAAAAGGTTCTCTGCAATCAATAGAGCTGAGAGCTGAGAGAG 3882
Db 481 CTTGCTAAAAAGAAAAGGTTCTCTGCAATCAATAGAGCTGAGAGCTGAGAGAG 540
QY 3883 GAGAAAGCTCGCTGTCAGAGCTAGAGGAAGCCCTTCAAGAAAGCCGATCGAGCTCCG 3942
Db 541 GAGAAAGCTCGCTGTCAGAGCTAGAGGAAGCCCTTCAAGAAAGCCGATCGAGCTCCG 600
QY 3943 TCCGCCCGGAGAGAGCTCCCAACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 4002
Db 601 TCCGCCCGGAGAGAGCTCCCAACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 660
QY 4003 GGCACCGCGAGGAGAGAGTGGCATGTCCGCATGTGCGGTGCGCAGAGCAAGAGCC 4062
Db 661 GGCACCGCGAGGAGAGAGTGGCATGTGCGGTGCGCAGAGCAAGAGCC 720
```

```
QY 4063 AGTCCATGAGCTGTGAGGCCCGCCATCCAGCCGAGAAAGAGTCTTAACCTCCAGAG 4122
Db 721 AGTCCATGAGCTGTGAGGCCCGCCATCCAGCCGAGAAAGAGTCTTAACCTCCAGAG 780
QY 4123 GAATTTAGTGGGCTCTTAAGAAAGCATGACCAACAAATATTTCTCCAGATTCAGCTA 4182
Db 781 GAATTTAGTGGGCTCTTAAGAAAGCATGACCAACAAATATTTCTCCAGATTCAGCTA 840
QY 4183 GAGCTGAACATGGAGGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4242
Db 841 GAGCTGAACATGGAGGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
QY 4243 CAGGATCCAAATGTCTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 4302
Db 901 CAGGATCCAAATGTCTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG 960
QY 4303 CCGAGCAGCTCGGCTTGGCTGCTGAATATGCAACACATTCACCGAGGCTTTCGCTG 4362
Db 961 CCGAGCAGCTCGGCTTGGCTGCTGAATATGCAACACATTCACCGAGGCTTTCGCTG 1020
QY 4363 GACAAATGAACTCCCGAGGCTCCAGACCAAGAGCCGAGCAGCTTGCACCTGGAA 4422
Db 1021 GACAAATGAACTCCCGAGGCTCCAGACCAAGAGCCGAGCAGCTTGCACCTGGAA 1080
QY 4423 GGGTGAATGAAGTGGCCAGAGATTAACAAAGAGCAGCAAGGCTGGGACAGAGATAC 4482
Db 1081 GGGTGAATGAAGTGGCCAGAGATTAACAAAGAGCAGCAAGGCTGGGACAGAGATAC 1140
QY 4483 ATTGCTCTGAGAGGATCAAAAAGTCTCATTTATGACATGAAGCCAGAGAGCTGACAG 4542
Db 1141 ATTGCTCTGAGAGGATCAAAAAGTCTCATTTATGACATGAAGCCAGAGAGCTGACAG 1200
QY 4543 AGCCGAGTGAAGAAATTTAGAGTGTGCTCCGAGGAGGAGATATATTCATGTGGCC 4602
Db 1201 AGCCGAGTGAAGAAATTTAGAGTGTGCTCCGAGGAGGAGATATATTCATGTGGCC 1260
QY 4603 GTTGTGCTTCCGAATCCGAATTAACAGCCAAAGAGATGTCCCATACATCTGAAGATG 4662
Db 1261 GTTGTGCTTCCGAATCCGAATTAACAGCCAAAGAGATGTCCCATACATCTGAAGATG 1320
QY 4663 GAATTCACCCGACACACCTGCTGCGCCGAGAAACCTCTGATCTGAGCTCCAGC 4722
Db 1321 GAATTCACCCGACACACCTGCTGCGCCGAGAAACCTCTGATCTGAGCTCCAGC 1380
QY 4723 TTCCCTGACAAACAGAGCTGGGCTCACCGCTTAAGATTCAGTGTCCAGGTTGGAGGTT 4782
Db 1381 TTCCCTGACAAACAGAGCTGGGCTCACCGCTTAAGATTCAGTGTCCAGGTTGGAGGTT 1440
QY 4783 TCTAGGAAAAAGCAGAGCTGATGCTAAATGCTTGAATCTCCCTGCTGAATCTGAA 4842
Db 1441 TCTAGGAAAAAGCAGAGCTGATGCTAAATGCTTGAATCTCCCTGCTGAATCTGAA 1500
QY 4843 GGTGATGACCGCTTGAACATGAATGCAAGCTGCTCTTCAAGTGAACAGAGTGTGGTG 4902
Db 1501 GGTGATGACCGCTTGAACATGAATGCAAGCTGCTCTTCAAGTGAACAGAGTGTGGTG 1560
QY 4903 GGCACCGAGAGAGGCTTACAGCCCTGAATGTCTTGAATAAATCCCTTAACCAATGCCA 4962
Db 1561 GGCACCGAGAGAGGCTTACAGCCCTGAATGTCTTGAATAAATCCCTTAACCAATGCCA 1620
QY 4963 GGAATTTGAGCAGTCTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
Db 1621 GGAATTTGAGCAGTCTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
QY 5023 GAGAGAGAGAGAGGAGCTGTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5082
Db 1681 GAGAGAGAGAGAGGAGCTGTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740
QY 5083 CAGTCCACCTGCTGCGCAGCCGACATCTCACCACCAATTTTGAAGCTGTCAAGGAG 5142
Db 1741 CAGTCCACCTGCTGCGCAGCCGACATCTCACCACCAATTTTGAAGCTGTCAAGGAG 1800
QY 5143 TGCCACTTGTGTTGGGCGAGGCAAGATTGAAGAGGAGCTGTGATCTGTGACAGCATGCC 5202
```

```

Db      1801  TGCCACTGTTGGGGGAGGAGATTGAGAACGGGCTGTGCATCTGTGAGCCATCCC 1860
QY      5203  AGCAAGTCGATCTTCCTCCGTACAGGAAAACCTGAGCAATCTGCATTCGGAAAGAG 5262
Db      1861  AGCAAGTCGATCTTCCTCCGTACAGGAAAACCTGAGCAATCTGCATTCGGAAAGAG 1920
QY      5263  ATAGAGCCTCAGAGCCCTGAGCGTATCACTTCACCAATTCATATCTCATTTGA 5322
Db      1921  ATAGAGCCTCAGAGCCCTGAGCGTATCACTTCACCAATTCATATCTCATTTGA 1980
QY      5323  ACCAATTAATTTCTAGAAATGACATGAGCAATGACAGCTCGAGAAATTCCTGATPAAG 5382
Db      1981  ACCAATTAATTTCTAGAAATGACATGAGCAATGAGCAATGAGCAATTCCTGATPAAG 2040
QY      5383  AATGACCATCTCTGGGACCTGCTGTGTTTGGCGCTCTTCACACACTTCCCTGTCTCA 5442
Db      2041  AATGACCATCTCTGGGACCTGCTGTGTTTGGCGCTCTTCACACACTTCCCTGTCTCA 2100
QY      5443  ATCGTGAGGTGAA CAGCGCAGGGCAGGAGAGAGTACTGTGTTCACGAATTT 5502
Db      2101  ATCGTGAGGTGAA CAGCGCAGGGCAGGAGAGAGTACTGTGTTCACGAATTT 2160
QY      5503  GAGGTGTCGTGATCTTCTTACGAGAAAGCTAGCCGCA CAGAGTCAAGTGAAGTCGC 5562
Db      2161  GAGGTGTCGTGATCTTCTTACGAGAAAGCTAGCCGCA CAGAGTCAAGTGAAGTCGC 2220
QY      5563  TTACCTTTGACCTTGTGCTTACAGAAACCTATCTGTTTGTGACCACTTCACTCACTC 5622
Db      2221  TTACCTTTGACCTTGTGCTTACAGAAACCTATCTGTTTGTGACCACTTCACTCACTC 2280
QY      5623  GAAATTAATTGATTCAGAGGACGCTCTCTCAGCAGGAAACCTGCCCCAGAGCTTCTGAC 5682
Db      2281  GAAATTAATTGATTCAGAGGACGCTCTCTCAGCAGGAAACCTGCCCCAGAGCTTCTGAC 2340
QY      5683  ATCCGGAACCCGCGTACCTGAGCCCTGCACTTTCTCTCAGAGAGGATTTACTTGGCTCC 5742
Db      2341  ATCCGGAACCCGCGTACCTGAGCCCTGCACTTTCTCTCAGAGAGGATTTACTTGGCTCC 2400
QY      5743  TCATACAGAGATTAATTAAAGGCTCATTTGTCTGCAAGGAAACCTGTGAAAGAGTCCGAC 5802
Db      2401  TCATACAGAGATTAATTAAAGGCTCATTTGTCTGCAAGGAAACCTGTGAAAGAGTCCGAC 2460
QY      5803  ACTGAACACCAACCCGCGGCGCTCCACCTCCCGGAGCAGCCCAACAGCAGGAGCCCAACC 5862
Db      2461  ACTGAACACCAACCCGCGGCGCTCCACCTCCCGGAGCAGCCCAACAGCAGGAGCCCAACC 2520
QY      5863  ACGTACCAAGCATCATCAACCAAGCGCTGACCTCAAGCCCAAGCGCCGCGGAGAGCCCC 5922
Db      2521  ACGTACCAAGCATCATCAACCAAGCGCTGACCTCAAGCCCAAGCGCCGCGGAGAGCCCC 2580
QY      5923  AGCCACCCGCGAGAGCCCAAGACACCCCAACCGCTTACCGCGAGGGGCGGACCTGAGTCGC 5982
Db      2581  AGCCACCCGCGAGAGCCCAAGACACCCCAACCGCTTACCGCGAGGGGCGGACCTGAGTCGC 2640
QY      5983  AGGGAACAAGTCTCTCGCGCGCCCTGTGAGGAGAGAGTCCCGCGGCGGATCTACAGC 6042
Db      2641  AGGGAACAAGTCTCTCGCGCGCCCTGTGAGGAGAGAGTCCCGCGGCGGATCTACAGC 2700
QY      6043  ACGCGAGAGAGCGGCTCCCGCGAGGCTGTTTGAAGACAGACCAAGAGGCGCGGCTGCTCT 6102
Db      2701  ACGCGAGAGAGCGGCTCCCGCGAGGCTGTTTGAAGACAGACCAAGAGGCGCGGCTGCTCT 2760
QY      6103  GCGGAGAGCGTGAAGAGCCCGCTGTCCAGAGTGAACAAGCTGAGGACCAAGTCTTCAAGTA 6162
Db      2761  GCGGAGAGCGTGAAGAGCCCGCTGTCCAGAGTGAACAAGCTGAGGACCAAGTCTTCAAGTA 2820
QY      6163  TAA 6165
Db      2821  TAA 2823

```

RESULT 14

```

ABV30132/c
ID      ABV30132 standard; cDNA; 2896 BP.
XX
AC      ABV30132;
XX
XX      16-SEP-2002 (first entry)
XX
DE      Human prostate expression marker cDNA 30123.
XX
KW      Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX      pharmacogenomic marker; gene; ss.
XX
OS      Homo sapiens.
XX
PN      M0200160860-A2.
XX
PD      23-AUG-2001.
XX
PF      20-FEB-2001; 2001WO-US05171.
XX
PR      17-FEB-2000; 2000US-183319P.
XX
PR      16-MAR-2000; 2000US-189862P.
XX
PR      25-MAY-2000; 2000US-207454P.
XX
PR      09-JUN-2000; 2000US-211314P.
XX
PR      18-JUL-2000; 2000US-219007P.
XX
PR      13-DEC-2000; 2000US-255281P.
XX
PA      (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI      Schlegel R, Endege WO, Monahan JB;
XX
XX      WPI, 2001-662795/76.
XX
PT      Novel isolated nucleic acid molecule associated with cancerous state of
XX      prostate cells and correlating with presence of prostate cancer, useful
XX      for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS      Claim 1, Page 6521-6522, 11750pp; English.
XX
XX
XX      The invention relates to an isolated nucleic acid molecule (1) comprising
XX      a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX      specification or its complement. (1) is useful for:
XX      (a) assessing whether a patient is afflicted with prostate cancer;
XX      (b) monitoring the progression of prostate cancer in a patient;
XX      (c) assessing the efficacy of a test compound to inhibit prostate
XX      cancer in a patient;
XX      (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX      in a patient;
XX      (e) selecting a composition for inhibiting prostate cancer in a patient;
XX      (f) assessing the prostate cell carcinogenic potential of a compound;
XX      (g) determining whether prostate cancer has metastasized in a patient;
XX      (h) assessing the aggressiveness or indolence of prostate cancer in a
XX      patient;
XX      (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ      Sequence 2896 BP; 525 A; 839 C; 647 G; 885 T; 0 other;
XX
XX
Query Match      43.5%; Score 2682.8; DB 23; Length 2896;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 2825; Conservative 0; Mismatches 2; Indels 111; Gaps 3;
QY      744  AAACAAGATGGAATGACCAACTCCGATTGGAGACCCAGATTTCATGAGCTCTGAACT 803
Db      2896  AAACAAGATGGAATGACCAACTCCGATTGGAGACCCAGATTTCATGAGCTCTGAACT 2837
QY      804  GCTGACTGTGATGAACGGGAGATGAAAAGACCTTACGGCTGAGACTGTGACTGTGATGTC 863
Db      2836  GCTGACTGTGATGAACGGGAGATGAAAAGGACCTTACGGCTGAGACTGTGATGATGATGTC 2777
QY      864  AGTGGCGGTGATGCTTATGAGATGATTTATGAGGAGATCCCTTCGAGAGGAACTTC 923
Db      2776  AGTGGCGGTGATGCTTATGAGATGATTTATGAGGAGATCCCTTCGAGAGGAACTTC 2717

```


Db 556 CACACACCAAACTTCTACTTGTCCAAACACATGATGAGGCTTCTGGCCGACACAGCA 497
 QY 3075 GATTGTACACTGCGAGAGTGAAGTGAACATCTCCGCGGAGATCAACGAGACGAGAT 3134
 Db 496 GATTGTACACTGCGAGAGTGAAGTGAACATCTCCGCGGAGATCAACGAGACGAGAT 437
 QY 3135 GCAGCTTACCGCCGAGAGCAACGATGAGGCTCTGAAGACCGAGCAACATGCTGGA 3194
 Db 436 GCAGCTTACCGCCGAGAGCAACGATGAGGCTCTGAAGACCGAGCAACATGCTGGA 377
 QY 3195 GGAACAGGTCATGATTTTGAAGGCTTAAACGATGAGCTGTAGAAAAAGAGCGGACGTG 3254
 Db 376 GGAACAGGTCATGATTTTGAAGGCTTAAACGATGAGCTGTAGAAAAAGAGCGGACGTG 317
 QY 3255 GGAAGCTTGAAGAGGCTCTCTGGTGTATGAGAAATCCAGTTTGAAGTGTCCGGTTCCAGA 3314
 Db 316 GGAAGCTTGAAGAGGCTCTCTGGTGTATGAGAAATCCAGTTTGAAGTGTCCGGTTCCAGA 257
 QY 3315 GCTGCAGAGATGTCTGACACCGAGAAACAGAGAGGCGAGAGCCGATCAGCGGATCAC 3374
 Db 256 GCTGCAGAGATGTCTGACACCGAGAAACAGAGAGGCGAGAGCCGATCAGCGGATCAC 197
 QY 3375 CGAGTCTGCCAGGTGTGTGAGCTGTGACGTGAAGAGACACAGGCTGTGATTTCTGCTCT 3434
 Db 196 CGAGTCTGCCAGGTGTGTGAGCTGTGACGTGAAGAGACACAGGCTGTGATTTCTGCTCT 137
 QY 3435 GCAGAGGCTCTCAAAAGAGCAGAAAGCTGAAGGCGCAGAGCTCTCTGCAAGCTCAATGA 3494
 Db 136 GCAGAGGCTCTCAAAAGAGCAGAAAGCTGAAGGCGCAGAGCTCTCTGCAAGCTCAATGA 77
 QY 3495 CCTGAGAGAAAGCATGCTATGCTTGAATAATGATGCCGAAGCTTACAGCAGAACTGGA 3554
 Db 76 CCTGAGAG-----AGAGCTGGA 59
 QY 3555 GACTGAAGAGAGCTCAACAGAGGCTCTGAAAGAGCAAGCCAAATTACAGCGACG 3612
 Db 58 GACTGAAGAGAGCTCAACAGAGGCTCTGAAAGAGCAAGCCAAATTACAGCGACG 1
 RESULT 15
 ABZ68725
 ID ABZ68725 standard; DNA; 1485 BP.
 AC ABZ68725;
 XX
 DT 16-MAY-2003 (first entry)
 XX
 DE Nucleotide sequence of citron rho/rac-interacting kinase-short kinase.
 KM Human; citron rho/rac-interacting kinase-short kinase; obesity;
 KM chronic obstructive pulmonary disease; hypertension; diabetes;
 KM coronary artery disease; hyperlipidaemia; stroke; gallbladder disease;
 KM gout; osteoarthritis; sleep apnea; cancer; thrombotic disease;
 KM polycystic ovarian syndrome; fertility; depression; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1485
 FT /*tag= a
 FT /partial
 FT /product= "citron rho/rac-interacting kinase-short
 FT kinase"
 XX
 XX WO2003004629-A2.
 XX
 XX 16-JAN-2003.
 XX
 XX 01-JUL-2002; 2002MO-EP07229.
 XX
 XX 02-JUL-2001; 2001US-301853P.
 XX 10-DEC-2001; 2001US-337130P.
 XX 25-APR-2002; 2002US-375015P.
 XX

XX (PAB) BAYER AG.
 PA
 XX
 PT Zhu Z;
 XX
 DR WPI: 2003-221595/21.
 DR P-PDB; ABP97681.
 XX
 PT New human citron rho/rac-interacting kinase-short kinase polypeptide
 PT and polynucleotide for preventing or treating diseases associated with
 PT the polypeptide dysfunction, e.g. obesity or chronic obstructive
 PT pulmonary disease
 PS
 PS Claim 1: Fig 1; 145pp; English.
 CC The present sequence encodes a human citron rho/rac-interacting
 CC kinase-short kinase polypeptide. The polynucleotide and polypeptide of
 CC the invention are useful in preventing, ameliorating, or treating
 CC diseases associated with the polypeptide dysfunction. The expression
 CC vector or the reagent is useful in the preparation of a medicament for
 CC modulating the activity of a human citron rho/rac-interacting
 CC kinase-short kinase in a disease, such as obesity or chronic obstructive
 CC pulmonary disease. These may also be used for treating obesity/
 CC overweight-associated comorbidities, such as hypertension, diabetes,
 CC coronary artery disease, hyperlipidaemia, stroke, gallbladder disease,
 CC gout, osteoarthritis, sleep apnea, cancer, thrombotic diseases,
 CC polycystic ovarian syndrome, reduced fertility, and depression.
 XX
 SQ Sequence 1485 BP; 373 A; 353 C; 372 G; 387 T; 0 other:
 Query Match 22.7%; Score 1401.4; DB 25; Length 1485;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATTGTAAAGTTCAAATATGAGAGCGGAAATCCTTTGAGATGCTGCTGTAACCCATT 60
 Db 1 ATTGTAAAGTTCAAATATGAGAGCGGAAATCCTTTGAGATGCTGCTGTAACCCATT 60
 QY 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCCAGGGAAACCACTTTATGACTCAA 120
 Db 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCCAGGGAAACCACTTTATGACTCAA 120
 QY 121 CAGCAGATGTCCTCTTTTCCGAGAGGAGATTTGATGCCCTCTTTGCTCTTTGAA 180
 Db 121 CAGCAGATGTCCTCTTTTCCGAGAGGAGATTTGATGCCCTCTTTGCTCTTTGAA 180
 QY 121 CAGCAGATGTCCTCTTTTCCGAGAGGAGATTTGATGCCCTCTTTGCTCTTTGAA 180
 Db 121 CAGCAGATGTCCTCTTTTCCGAGAGGAGATTTGATGCCCTCTTTGCTCTTTGAA 180
 QY 181 GAATGAGTCAGGCTGCTCTGATGAAGATTAAAGCAGTGAACAATTGTCCGAGATAT 240
 Db 181 GAATGAGTCAGGCTGCTCTGATGAAGATTAAAGCAGTGAACAATTGTCCGAGATAT 240
 QY 241 TCCGACACCATAGCTAGTTTACAGAGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300
 Db 241 TCCGACACCATAGCTAGTTTACAGAGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300
 QY 241 TCCGACACCATAGCTAGTTTACAGAGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300
 Db 241 TCCGACACCATAGCTAGTTTACAGAGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300
 QY 301 AGTCTGTGAGTGTGTGTGCTACCTTGTGAAGGACAGTGTGAAGAGAAAGCAACGG 360
 Db 301 AGTCTGTGAGTGTGTGTGCTACCTTGTGAAGGACAGTGTGAAGAGAAAGCAACGG 360
 QY 301 AGTCTGTGAGTGTGTGTGCTACCTTGTGAAGGACAGTGTGAAGAGAAAGCAACGG 360
 Db 301 AGTCTGTGAGTGTGTGTGCTACCTTGTGAAGGACAGTGTGAAGAGAAAGCAACGG 360
 QY 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATTTGGCCAGAGACAGTTTCA 420
 Db 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATTTGGCCAGAGACAGTTTCA 420
 QY 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATTTGGCCAGAGACAGTTTCA 420
 Db 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATTTGGCCAGAGACAGTTTCA 420
 QY 421 TTTTGGAGAGAGGAGGAGCAATATATCTCCAGAGCAAGCCGTGATCCCAATTA 480
 Db 421 TTTTGGAGAGAGGAGGAGCAATATATCTCCAGAGCAAGCCGTGATCCCAATTA 480
 QY 481 CAGTATGCTTTCCAGAGCAAAATCACTTTATCTGTGATGAGATATAGCTGAGGG 540
 Db 481 CAGTATGCTTTCCAGAGCAAAATCACTTTATCTGTGATGAGATATAGCTGAGGG 540
 QY 541 GACTTGCTGTCACTTTGATATGATATGAGACCAATTGATGAAACCTGATACGTTT 600
 Db 541 GACTTGCTGTCACTTTGATATGATATGAGACCAATTGATGAAACCTGATACGTTT 600

QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAACAGCGTTCTGATGGGATACGTGCATCGA 660
Db 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAACAGCGTTCTGATGGGATACGTGCATCGA 660
QY 661 GACATCAAGCCTGAGACATTTCTGTTGACCGACAGAGACATCAAGCTGGTGAATTT 720
Db 661 GACATCAAGCCTGAGACATTTCTGTTGACCGACAGAGACATCAAGCTGGTGAATTT 720
QY 721 GGATCTGCCCGGAAAAATGAATTCAACCAAGTGGTGAATGCCAACTCCCGATTGGGACC 780
Db 721 GGATCTGCCCGGAAAAATGAATTCAACCAAGTGGTGAATGCCAACTCCCGATTGGGACC 780
QY 781 CCAATTAACATGGCTCCTGTAAGTGTGATGTAACGGGATGGAAGGACCTAC 840
Db 781 CCAATTAACATGGCTCCTGTAAGTGTGATGTAACGGGATGGAAGGACCTAC 840
QY 841 GGCTGGACTGTGACTGTGGTCACTGAGTGGCGGTGATGCTTAATGATGATTTATGGGAGA 900
Db 841 GGCTGGACTGTGACTGTGGTCACTGAGTGGCGGTGATGCTTAATGATGATTTATGGGAGA 900
QY 901 TCCCGCTTCCGAGAGGGGAACTCTGCCGAACTTCAATTAATGATTTCCAGCGG 960
Db 901 TCCCGCTTCCGAGAGGGGAACTCTGCCGAACTTCAATTAATGATTTCCAGCGG 960
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGTGATCTTGAATGATTTCAAGC 1020
Db 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGTGATCTTGAATGATTTCAAGC 1020
QY 1021 TTGTTGGCGGCCGAGAAAGAGAGTGAAGTTGAAGTCTTGGCTGCCATCCTTTCTTC 1080
Db 1021 TTGTTGGCGGCCGAGAAAGAGAGTGAAGTTGAAGTCTTGGCTGCCATCCTTTCTTC 1080
QY 1081 TCTAAATTTGACTGGAACAACATTCGTAACCTCTCCCGCTCGTTGCCACCGTCAAG 1140
Db 1081 TCTAAATTTGACTGGAACAACATTCGTAACCTCTCCCGCTCGTTGCCACCGTCAAG 1140
QY 1141 TCTGACGATGACACCTCCATTTTGTATGAACACAGAGAAATTCGTGGTTCATCCTCT 1200
Db 1141 TCTGACGATGACACCTCCATTTTGTATGAACACAGAGAAATTCGTGGTTCATCCTCT 1200
QY 1201 CCGTGCAAGCTGAGCGCCCTCAAGGCTCTCGGGTGAAGAACTGCCGTTTGGGGTTTGG 1260
Db 1201 CCGTGCAAGCTGAGCGCCCTCAAGGCTCTCGGGTGAAGAACTGCCGTTTGGGGTTTGG 1260
QY 1261 TACAGCAAGGCACTGGGGATTTCTGTAGATCTGAGTCTGTGTGTGGGTCTGAACTCC 1320
Db 1261 TACAGCAAGGCACTGGGGATTTCTGTAGATCTGAGTCTGTGTGTGGGTCTGAACTCC 1320
QY 1321 CCTGCCAAGACTAGCTCATGGAAGAAATTTCTCATCAAAAGCAAGCTTCAAGAC 1380
Db 1321 CCTGCCAAGACTAGCTCATGGAAGAAATTTCTCATCAAAAGCAAGCTTCAAGAC 1380
QY 1381 TCTCAGGCAAGTGTCAAGAT 1403
Db 1381 TCTCAGGCAAGTGTCAAGAT 1403

Search completed: November 15, 2003, 04:22:56
Job time : 989.601 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 01:08:15, Search time 14461.8 Seconds
(without alignments)
17439.597 Million cell updates/sec

Title: US-10-028-946-1

Perfect score: 6165

Sequence: 1 atgttgagttcaaatatg.....ggaccagctcttcagratataa 6165

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_ov.*
7: gb_ov.*
8: gb_ov.*
9: gb_ov.*
10: gb_ov.*
11: gb_ov.*
12: gb_ov.*
13: gb_ov.*
14: gb_ov.*
15: gb_ov.*
16: gb_ov.*
17: gb_ov.*
18: gb_ov.*
19: gb_ov.*
20: gb_ov.*
21: gb_ov.*
22: gb_ov.*
23: gb_ov.*
24: gb_ov.*
25: gb_ov.*
26: gb_ov.*
27: gb_ov.*
28: gb_ov.*
29: gb_ov.*
30: gb_ov.*
31: gb_ov.*
32: gb_ov.*
33: gb_ov.*
34: gb_ov.*
35: gb_ov.*
36: gb_ov.*
37: gb_ov.*
38: gb_ov.*
39: gb_ov.*
40: gb_ov.*
41: gb_ov.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6165	100.0	6165	AX574425	AX574425 Sequence
2	6161.8	99.9	6298	AX504254	AX504254 Sequence
3	6160.2	99.9	6165	AX671105	AX671105 Sequence
4	6160.2	99.9	8603	AX671108	AX671108 Sequence
5	6142.2	99.6	6156	AX671112	AX671112 Sequence
6	6126	99.4	6159	AX166510	AX166510 Sequence
7	6121.8	99.3	6189	AX503780	AX503780 Sequence
8	6105.4	99.0	6201	AX503778	AX503778 Sequence
9	5843.2	94.8	8576	AX257469	AX257469 Homo sapi
10	5837	94.7	5877	AX574427	AX574427 Sequence
11	5661.4	91.8	6159	AX429514	AX429514 Sequence
12	5661.4	91.8	6574	AX429512	AX429512 Sequence
13	5105.2	82.8	6554	AX068824	AX068824 Mus muscu
14	4021.2	65.2	4967	AF070066	AF070066 Mus muscu
15	4018.2	65.2	5019	MM033904	U99904 Mus musculu
16	3906.6	63.4	5952	AF039218	AF039218 Rattus no
17	2818.2	45.7	5261	AX671110	AX671110 Sequence
18	2818.2	45.7	5261	AB023166	AB023166 Homo sapi
19	2403	39.0	3495	BC051165	BC051165 Mus muscu
20	2403	39.0	3506	BC023775	BC023775 Mus muscu
21	1401.4	22.7	1485	AX671037	AX671037 Sequence
22	1401.4	22.7	1785	AX671044	AX671044 Sequence
23	1399.8	22.7	2066	AX642956	AX642956 Sequence
24	1398.2	22.7	1515	AR253937	AR253937 Sequence
25	1088.8	17.7	2380	AX671042	AX671042 Sequence
26	1088.8	17.7	2380	AF086823	AF086823 Mus muscu
27	1080.8	17.5	1386	AF070065	AF070065 Rattus no
28	754	12.2	1058	AX053315	AX053315 Sequence
29	698	11.3	817	AX053416	AX053416 Sequence
30	350.4	5.7	354	BD059649	BD059649 Secreted
31	311.2	5.0	136436	AC002563	AC002563 Human PAC
32	311.2	5.0	147862	AC079317	AC079317 Homo sapi
33	311.2	5.0	190514	AC026363	AC026363 Homo sapi
34	311.2	5.0	208686	AC079262	AC079262 Homo sapi
35	258	4.2	258	AR269950	AR269950 Sequence
36	257.2	4.2	207404	AC115795	AC115795 Mus muscu
37	256.2	4.2	32658	AC135099	AC135099 Rattus no
38	256.2	4.2	234312	AC133614	AC133614 Rattus no
39	256.2	4.2	248557	AC097432	AC097432 Rattus no
40	250	4.1	254	BD071469	BD071469 Secreted
41	246.8	4.0	1530	AX504235	AX504235 Sequence
42	241.6	3.9	6335	AF021935	AF021935 Rattus no
43	235.2	3.8	4917	AY166586	AY166586 Homo sapi
44	233	3.8	2785	HS059505	U59305 Human ser-t
45	233	3.8	4917	HSN518976	JS18976 Homo sapi

ALIGNMENTS

RESULT 1
AX574425
LOCUS AX574425 6165 bp DNA linear PAT 07-JAN-2003
DEFINITION Sequence 1 from Patent WO02059325.
ACCESSION AX574425
VERSION AX574425.1 GI:27551751
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Yu, X., Miranda, M. and Fridale, C.J.
TITLE Human kinases and polynucleotides encoding the same
JOURNAL Patent: WO 02059325-A 1 01-AUG-2002;

QY	901	TTCCCTTCGAGAGGAACTCTCCGCAAACTTCATTAACATTATGAATTTCCAGCG	960
Db	901	TTCCCTTCGAGAGGAACTCTCCGCAAACTTCATTAACATTATGAATTTCCAGCG	960
QY	961	TTTTTGAATTTCCAGATGACCCCAAGAGGAGTGACTTTCTTGATCTGATTCAAAGC	1020
Db	961	TTTTTGAATTTCCAGATGACCCCAAGAGGAGTGACTTTCTTGATCTGATTCAAAGC	1020
QY	1021	TTGTTGTGCGCGCAGAAAAGAGACTGAAGTTTGAAGCTCTTGTGCTCATCTTTCTTC	1080
Db	1021	TTGTTGTGCGCGCAGAAAAGAGACTGAAGTTTGAAGCTCTTGTGCTCATCTTTCTTC	1080
QY	1081	TTCTAAATTGAGTGGAAACAATTGTGTAACCTCTCTCCCTTGTGTTCCACCTCAAG	1140
Db	1081	TTCTAAATTGAGTGGAAACAATTGTGTAACCTCTCTCCCTTGTGTTCCACCTCAAG	1140
QY	1141	TTCTGACGATGACACCTCCATTTTTGATGAACAGAGAAAGAAATTTGGGGTTCAATCCCT	1200
Db	1141	TTCTGACGATGACACCTCCATTTTTGATGAACAGAGAAAGAAATTTGGGGTTCAATCCCT	1200
QY	1201	CCGTGCGAGCTGAGGCCCCCTCAGGCTTCTCGGGTAAAGAACTGCCGTTTGTGGGGTTTTCG	1260
Db	1201	CCGTGCGAGCTGAGGCCCCCTCAGGCTTCTCGGGTAAAGAACTGCCGTTTGTGGGGTTTTCG	1260
QY	1261	TACAGCAAGGACTGGGGATCTTGTTGATGACTGAGCTGTGTTGTGGGGCTGAGACTCC	1320
Db	1261	TACAGCAAGGACTGGGGATCTTGTTGATGACTGAGCTGTGTTGTGGGGCTGAGACTCC	1320
QY	1321	CTGCGCAAGACTAGCTCCATGAGAAAGAAACCTTCATCAAAAGCAAGACTTACAAGAC	1380
Db	1321	CTGCGCAAGACTAGCTCCATGAGAAAGAAACCTTCATCAAAAGCAAGACTTACAAGAC	1380
QY	1381	TTCTCAGGACAGTGTCACAGATGGAGTGGACAGAAATGACCCGCTTACATCGAGAGTGTCA	1440
Db	1381	TTCTCAGGACAGTGTCACAGATGGAGTGGACAGAAATGACCCGCTTACATCGAGAGTGTCA	1440
QY	1441	GAGGTGAGGCTGTGCTTGAATGTCAGAGAGAGTGGAGCTGAAGGCTCTGAGACTCAAGA	1500
Db	1441	GAGGTGAGGCTGTGCTTGAATGTCAGAGAGAGTGGAGCTGAAGGCTCTGAGACTCAAGA	1500
QY	1501	TTCCCTCTCTGAGAGCGAGACTTGTCTACTACATCAAGAAATGCAATAGCTTTAAAGCGAAGT	1560
Db	1501	TTCCCTCTCTGAGAGCGAGACTTGTCTACTACATCAAGAAATGCAATAGCTTTAAAGCGAAGT	1560
QY	1561	TTTGAGCAGAAGCAGATGAGAGGTGTCTCCAGAGAGTGAACAACACTGCGACTTCCCAT	1620
Db	1561	TTTGAGCAGAAGCAGATGAGAGGTGTCTCCAGAGAGTGAACAACACTGCGACTTCCCAT	1620
QY	1621	GATATCAAGAGCAGAGCCGGAAGCTTCCAAAGAAATCAAAAGAGCAGAGTTACAGAGCTCAA	1680
Db	1621	GATATCAAGAGCAGAGCCGGAAGCTTCCAAAGAAATCAAAAGAGCAGAGTTACAGAGCTCAA	1680
QY	1681	GTGGAAAGAAATGAGGTGTGAATGATCACTAGTGGAGAGAGATTTGTCTCAGCAGAAAGA	1740
Db	1681	GTGGAAAGAAATGAGGTGTGAATGATCACTAGTGGAGAGAGATTTGTCTCAGCAGAAAGA	1740
QY	1741	CGAGTGAATCTTACGATCTGAGACTGAGAGAGTCTCGGCTTGCTGCTGAAGAAATTCAG	1800
Db	1741	CGAGTGAATCTTACGATCTGAGACTGAGAGAGTCTCGGCTTGCTGCTGAAGAAATTCAG	1800
QY	1801	CGAAGAGCGACAAATGTCAGCTTAACTGTTGAAAGGCTTAAGGATCAAGGGAAGGCTGAA	1860
Db	1801	CGAAGAGCGACAAATGTCAGCTTAACTGTTGAAAGGCTTAAGGATCAAGGGAAGGCTGAA	1860
QY	1861	GTGGAGAAATATCGAAACTGAGAAAGATCAATGCTGAGCAGACAGTCAAAATTCAGAG	1920
Db	1861	GTGGAGAAATATCGAAACTGAGAAAGATCAATGCTGAGCAGACAGTCAAAATTCAGAG	1920
QY	1921	CTTCAAGAGAAACTGAGAAAGGCTGTAAAGCAGACACGAGAGCCAACGAGCTGTGCGAG	1980
Db	1921	CTTCAAGAGAAACTGAGAAAGGCTGTAAAGCAGACACGAGAGCCAACGAGCTGTGCGAG	1980
QY	1981	AATATCCGACAGCAAGAGAGGAGCCTCGAGAGAGGAGCTGAGAAAGTCTGAACTTCAAG	2040
Db	1981	AATATCCGACAGCAAGAGAGGAGCCTCGAGAGAGGAGCTGAGAAAGTCTGAACTTCAAG	2040

QY	4201	ACAAAGTGTGCTGTGTGTCTTGGATACCGTGCACTTTTGACGCGCAGCATCCAAATGTCTC	4260
Db	4201	ACAAAGTGTGCTGTGTGTCTTGGATACCGTGCACTTTTGACGCGCAGCATCCAAATGTCTC	4260
QY	4261	GAATGTACAGTATGTGTACCCCAAGTGTCCAGTGTTCACGCCAACCCTGGCGCTTG	4320
Db	4261	GAATGTACAGTATGTGTACCCCAAGTGTGTCCAGTGTTCACGCCAACCCTGGCGCTTG	4320
QY	4321	CTGTGAAATATGCCACACACTTACCGAGGCTTTCTGCGTGAACAATATGAATCTCCCA	4380
Db	4321	CTGTGAAATATGCCACACACTTACCGAGGCTTTCTGCGTGAACAATATGAATCTCCCA	4380
QY	4381	GCTCTCCAGACCAAGAGCCCAAGCACTTGGACCTTGAAAGGCTGATGAAGTGTGCC	4440
Db	4381	GCTCTCCAGACCAAGAGCCCAAGCACTTGGACCTTGAAAGGCTGATGAAGTGTGCC	4440
QY	4441	AGGAATACAAACGAGGACAGCAAGGCTGGACAGAAAGTACATGTGCTGAGAGATCA	4500
Db	4441	AGGAATACAAACGAGGACAGCAAGGCTGGACAGAAAGTACATGTGCTGAGAGATCA	4500
QY	4501	AAAGTCTCATTTATATGACATGAGCCAGAGAGAGCTGGAACAGAGGCTGTGAAAGAAATTT	4560
Db	4501	AAAGTCTCATTTATATGACATGAGCCAGAGAGAGCTGGAACAGAGGCTGTGAAAGAAATTT	4560
QY	4561	GAGCTGTGCTTCCGACCGGGATGATATCTATTCATGTGAGCGTGTGGTCTCCGAATC	4620
Db	4561	GAGCTGTGCTTCCGACCGGGATGATATCTATTCATGTGAGCGTGTGGTCTCCGAATC	4620
QY	4621	GCAATATACGCCAAAGACAGATGTCCCATATCAATCTGAAAGATGAAATCTCACCCGACACC	4680
Db	4621	GCAATATACGCCAAAGACAGATGTCCCATATCAATCTGAAAGATGAAATCTCACCCGACACC	4680
QY	4681	ACCTGTGGCTCCGGAGAACCCCTACTGTGTAGTCCAGCTCCAGCTTCCCTGACAAACAGCC	4740
Db	4681	ACCTGTGGCTCCGGAGAACCCCTACTGTGTAGTCCAGCTCCAGCTTCCCTGACAAACAGCC	4740
QY	4741	TGGGTACCCGCTTAGAATCAGATTGTGCAGAGTGGAGAGTTTCTAGGAGAAAAAGCAGAA	4800
Db	4741	TGGGTACCCGCTTAGAATCAGATTGTGCAGAGTGGAGAGTTTCTAGGAGAAAAAGCAGAA	4800
QY	4801	GCTGATGTCTTAAACCTGCTTGAAACCTCCTGCTGAACTGGAAGGTGATGACCCGCTTAGAC	4860
Db	4801	GCTGATGTCTTAAACCTGCTTGAAACCTCCTGCTGAACTGGAAGGTGATGACCCGCTTAGAC	4860
QY	4861	ATGAATCAGACGCTCCCTTCAGTGAACAGAGTGGTGTGGAGGACCGAGGAAAGGCGTC	4920
Db	4861	ATGAATCAGACGCTCCCTTCAGTGAACAGAGTGGTGTGGAGGACCGAGGAAAGGCGTC	4920
QY	4921	TACGCCCTGATGTCTTTGAAAAACTCCCTAACCCATGTCCAGGAATTTGAGACAGTCTTC	4980
Db	4921	TACGCCCTGATGTCTTTGAAAAACTCCCTAACCCATGTCCAGGAATTTGAGACAGTCTTC	4980
QY	4981	CAAAATTTATATTAACAAGACCTTGAGAAAGTAACTCATGATAGCAGAGAAAGACCGAGCA	5040
Db	4981	CAAAATTTATATTAACAAGACCTTGAGAAAGTAACTCATGATAGCAGAGAAAGACCGAGCA	5040
QY	5041	CTGTGTCTTGTGACGTTGAAGAAAGTAAACAGTCCCTGGCCCAAGTCCCACTGCTGTGCC	5100
Db	5041	CTGTGTCTTGTGACGTTGAAGAAAGTAAACAGTCCCTGGCCCAAGTCCCACTGCTGTGCC	5100
QY	5101	CAGCCCGACATCTACCCCAATTTTGGAGCTGTCAAGGCGTGGCACTTTGTTGGGGACA	5160
Db	5101	CAGCCCGACATCTACCCCAATTTTGGAGCTGTCAAGGCGTGGCACTTTGTTGGGGACA	5160
QY	5161	GGCAAGATTGAAACGGGCTCTGTGATCTGTGACAGCATGCGCCAGCAAGTGTGATCTTC	5220
Db	5161	GGCAAGATTGAAACGGGCTCTGTGATCTGTGACAGCATGCGCCAGCAAGTGTGATCTTC	5220
QY	5221	CGCTACACAGAAACCTCAGCAAAATATGTGATCCGAGAAAGATAGACCTTGAGAGCC	5280
Db	5221	CGCTACACAGAAACCTCAGCAAAATATGTGATCCGAGAAAGATAGACCTTGAGAGCC	5280

QY	5281	TGCACCTGATCATCACTTACACCAATTAAGTATTCCTAATTTGAAACCAATAAATTCACGAA	5340
Dp	5281	TGCACCTGATCATCACTTACACCAATTAAGTATTCCTAATTTGAAACCAATAAATTCACGAA	5340
QY	5341	ATCGACATGAAAGCACTATACACGCTCGAGGAATTCCTGATATAGAAATGACATTCCTTGACA	5400
Dp	5341	ATCGACATGAAAGCACTATACACGCTCGAGGAATTCCTGATATAGAAATGACATTCCTTGACA	5400
QY	5401	CCTGCTGTGTTTGGCCGCTCTTCCAAACGCTTCCCTGTCTCAATGTGTACAGTGAACAGC	5460
Dp	5401	CCTGCTGTGTTTGGCCGCTCTTCCAAACGCTTCCCTGTCTCAATGTGTGAAGTGAACAGC	5460
QY	5461	GCAGGGACGACGAGAGAGATACCTGTGCTGTTTCCACGAAATTTGAGAGTTGTGTGGATTCT	5520
Dp	5461	GCAGGGACGACGAGAGAGATACCTGTGCTGTTTCCACGAAATTTGAGAGTTGTGTGGATTCT	5520
QY	5521	TACGGAAAGCTGATCCGCGACAGACGATCTCAAGTGAAGTTCGCTTACCTTTTGACCTTTGCC	5580
Dp	5521	TACGGAAAGCTGATCCGCGACAGACGATCTCAAGTGAAGTTCGCTTACCTTTTGACCTTTGCC	5580
QY	5581	TACAGAGAACCCCTATCTGTTTGTATGCCACATTCGAACCTCACTCGAAGATTAATGAAATCCAG	5640
Dp	5581	TACAGAGAACCCCTATCTGTTTGTATGCCACATTCGAACCTCACTCGAAGATTAATGAAATCCAG	5640
QY	5641	GCAGGCTCTTGACAGAGGAGACCCCTGCCGAGCGTACTGTGACATCCGAAACCCGCGCTAC	5700
Dp	5641	GCAGGCTCTTGACAGAGGAGACCCCTGCCGAGCGTACTGTGACATCCGAAACCCGCGCTAC	5700
QY	5701	CTGGGGCCCTGCAATTTCTCTAGAGAGGCAATTACTTGGCGCTCTATATACAGATTAATTA	5760
Dp	5701	CTGGGGCCCTGCAATTTCTCTAGAGAGGCAATTACTTGGCGCTCTATATACAGATTAATTA	5760
QY	5761	AGGGTCAATTTGCTGCAAGGAGAAACCTGTGTAAGAGATCCGAGCACTGAACACCAACCGGGGC	5820
Dp	5761	AGGGTCAATTTGCTGCAAGGAGAAACCTGTGTAAGAGATCCGAGCACTGAACACCAACCGGGGC	5820
QY	5821	CCGTCACACTCCCGACACACCTCCCAACACGAGGCTCAATCCAGTCAACAGACATCT	5880
Dp	5821	CCGTCACACTCCCGACACACCTCCCAACACGAGGCTCAATCCAGTCAACAGACATCT	5880
QY	5881	ACCAAGCGGTGAGCTTCAAGCCGACGCGCCGCGAGAGGCCCAACCCGCGAGAGGCCA	5940
Dp	5881	ACCAAGCGGTGAGCTTCAAGCCGACGCGCGCCGCGAGAGGCCCAACCCGCGAGAGGCCA	5940
QY	5941	AGCACAATCCCAACCGCTACCGCGAGGGGCGGACCGGAGCTGTGCGAGGACAAATCTCTGTGC	6000
Dp	5941	AGCACAATCCCAACCGCTACCGCGAGGGGCGGACCGGAGCTGTGCGAGGACAAATCTCTGTGC	6000
QY	6001	CGCCCCCTGAGAGGAGAGAGTCCCCCGGCGGATATCTACAGACGCGGAGAGAGCGGTCC	6060
Dp	6001	CGCCCCCTGAGAGGAGAGAGTCCCCCGGCGGATATCTACAGACGCGGAGAGAGCGGTCC	6060
QY	6061	CCGCGAGGCTGTTTGAAGACAGACAGACAGGGGCGGCTGTGTCGCGGAGCCGTGAAGACC	6120
Dp	6061	CCGCGAGGCTGTTTGAAGACAGACAGAGGAGCGGCTGTGTCGCGGAGCCGTGAAGACC	6120
QY	6121	CCGCTGTCCAGGATGAACAAGGCTGTGGACACAGTCTTACAGTATTA	6165
Dp	6121	CCGCTGTCCAGGATGAACAAGGCTGTGGACACAGTCTTACAGTATTA	6165

RESULT 2					
AX504254	AM504254	6298 bp	DNA	linear	PAT 27-SEP-2002D
LOCUS					
DEFINITION	Sequence 43 from Patent WO0233099.				
ACCESSION	AX504254				
VERSION	AX504254.1				
KEYWORDS	GI:2386094				
SOURCE	.				
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1				

AUTHORS

Gururajan, R., Baughn, M.R., Wajia, N.K., Elliott, V.S., Xu, Y.,
Aryizu, C., Yao, M.G., Ramkumar, J., Ding, L., Tang, Y.T., Hafila, A.J.,
Nguyen, D.B., Gandhi, A.R., Lu, Y., Yue, H., Burford, N., Bandman, O.,
Tribouley, C.M., Lal, P.G., Recipon, S.A., Lu, D.A., Borowsky, M.L.,
Thornlon, M., Swartner, A., Thangavelu, K., Khan, F.A. and Ison, C.H.
Human Kinases
Patent: WO 023099-A 43 25-APR-2002;

JOURNAL

Incyte Genomics, Inc. (US)

FEATURES

Location/Qualifiers

1..6298
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7484498CB1"

BASE COUNT 1772 a 1585 c 1720 g 1221 t

ORIGIN

Query Match 99.9%; Score 6161.8; DB 6; Length 6298;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGTTGAAGTCAATATATGAGCGGGAATCCTTTGATGCTGATGCTGAAACCAT 60
55 ATGTTGAAGTCAATATATGAGCGGGAATCCTTTGATGCTGATGCTGAAACCAT 114
61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCAGGGGAAACCACTTTATGACTCA 120
115 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCAGGGGAAACCACTTTATGACTCA 174
121 CAGAGAGTGTCTCTCTTCCCGAGAGGATATAGATCCCTCTTGTCTCTTGA 180
175 CAGAGAGTGTCTCTCTTCCCGAGAGGATATAGATCCCTCTTGTCTCTTGA 234
181 GAATGAGTCAAGCTGCTCTGATGAGATTAAAGCACTGAGCACTTTGTCCGGAAT 240
235 GAATGAGTCAAGCTGCTCTGATGAGATTAAAGCACTGAGCACTTTGTCCGGAAT 294
241 TCCGACACCAATAGCTGATGAGTAAAGGAGCTCCAGGCTCCGCAAGAGCTCCAG 300
295 TCCGACACCAATAGCTGATGAGTAAAGGAGCTCCAGGCTCCGCAAGAGCTCCAG 354
301 AGCTTGTAGGTTGTGCTCACTTGTCTGAGAGTGTGTAAGAGGAAAGCAACGG 360
355 AGCTTGTAGGTTGTGCTCACTTGTCTGAGAGTGTGTAAGAGGAAAGCAACGG 414
361 GACATCTATGCTATGAGAGTGAAGAGAGAGCTTTATTGGCCCAAGAGCTTTCA 420
415 GACATCTATGCTATGAGAGTGAAGAGAGAGCTTTATTGGCCCAAGAGCTTTCA 474
421 TTTTGAAGAGAGCGGAACATTTATCTGAGAGCAAGCCCTGAGATCCCCCAAT 480
475 TTTTGAAGAGAGCGGAACATTTATCTGAGAGCAAGCCCTGAGATCCCCCAAT 534
481 CAGTATGCTTTTCAAGCAAAATCACCTTTATCTGAGAGCAAGCCCTGAGAG 540
535 CAGTATGCTTTTCAAGCAAAATCACCTTTATCTGAGAGCAAGCCCTGAGAG 594
541 GACTTGTGCTCACTTTGAATAGATGAGAGCAAGTAAAGAAACCTGATACGTT 600
595 GACTTGTGCTCACTTTGAATAGATGAGAGCAAGTAAAGAAACCTGATACGTT 654
601 TACTTGTGCTCACTTTGAATAGATGAGAGCAAGTAAAGAAACCTGATACGTT 660
655 TACTTGTGCTCACTTTGAATAGATGAGAGCAAGTAAAGAAACCTGATACGTT 714
661 GACATCAAGCTGAGAACTTTCTGTTGAGAGCAAGCACTCAAGCTGAGATTT 720
715 GACATCAAGCTGAGAACTTTCTGTTGAGAGCAAGCACTCAAGCTGAGATTT 774
721 GATCTGCTCCGGAATGAATTCACACAGATGATGATGCTCCGATTTGGAGC 780
775 GATCTGCTCCGGAATGAATTCACACAGATGATGATGCTCCGATTTGGAGC 834

781 CCAGATTAATAGGCTCTGAAGTGTGATGATGAACGGGGAATGAAAAAGCACTAC 840
835 CCAAGTTTACATGCTCTGAACTGTGATGATGAACGGGGAATGAAAAAGCACTAC 894
841 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
895 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 954
901 TCCCTCTGAGAGGAACTCTGAGAGGAACTCTGAGAGGAACTCTGAGAGGAA 960
955 TCCCTCTGAGAGGAACTCTGAGAGGAACTCTGAGAGGAACTCTGAGAGGAA 1014
961 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020
1015 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1074
1021 TTTTGTGAGGAGCAAGAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1080
1075 TTTTGTGAGGAGCAAGAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1134
1081 TCTAAATTTGAGTGAAGCAACCTTGTGATGATGATGATGATGATGATGAT 1140
1135 TCTAAATTTGAGTGAAGCAACCTTGTGATGATGATGATGATGATGATGAT 1194
1141 TCTGAGATGAGCACTTCAATTTTGAATGAACCAAGAAATTTGAGTGAAGT 1200
1195 TCTGAGATGAGCACTTCAATTTTGAATGAACCAAGAAATTTGAGTGAAGT 1254
1201 CCGTCCAGCTGAGAGCCCTCAAGGCTTCTGAGTGAAGAGTCCGTTTGTGG 1260
1255 CCGTCCAGCTGAGAGCCCTCAAGGCTTCTGAGTGAAGAGTCCGTTTGTGG 1314
1261 TACAGCAAGCACTGAGGATTTCTGATGATGATGATGATGATGATGATGAT 1320
1315 TACAGCAAGCACTGAGGATTTCTGATGATGATGATGATGATGATGATGAT 1374
1321 CCGTCCAGCTGAGAGCCCTCAAGGCTTCTGAGTGAAGAGTCCGTTTGTGG 1380
1375 CCGTCCAGCTGAGAGCCCTCAAGGCTTCTGAGTGAAGAGTCCGTTTGTGG 1434
1381 TCTGAGCAAGTGTCAAAATGAGAGAGAGAGTCCGTTTGTGATGATGATGAT 1440
1435 TCTGAGCAAGTGTCAAAATGAGAGAGAGAGTCCGTTTGTGATGATGATGAT 1494
1441 GAGTGTGAGGCTGTGCTTGAATGAGAGAGTGAAGTGAAGTGAAGTGAAGT 1500
1495 GAGTGTGAGGCTGTGCTTGAATGAGAGAGTGAAGTGAAGTGAAGTGAAGT 1554
1501 TCCCTCTGAGAGGAACTTGTGATGATGATGATGATGATGATGATGATGAT 1560
1555 TCCCTCTGAGAGGAACTTGTGATGATGATGATGATGATGATGATGATGAT 1614
1561 TTTGAGCAAGCAAGGATGAGGATTTCCAGAGAGTGAAGCAAGCACTGAT 1620
1615 TTTGAGCAAGCAAGGATGAGGATTTCCAGAGAGTGAAGCAAGCACTGAT 1674
1621 GATATCAAGAGAGAGGAGGAGTCCCAAAATTCAGAAAGAGAGTGAAGTGAAG 1680
1675 GATATCAAGAGAGAGGAGGAGTCCCAAAATTCAGAAAGAGAGTGAAGTGAAG 1734
1681 GTGAGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGAT 1740
1735 GTGAGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGAT 1794
1741 CGAGTGAATCTTCAAGATCTGAGCTGAGAGAGTCTGGCTTGTCTGAGAAAT 1800
1795 CGAGTGAATCTTCAAGATCTGAGCTGAGAGAGTCTGGCTTGTCTGAGAAAT 1854
1801 CGAGAGAGAGAGAGTCAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1860
1855 CGAGAGAGAGAGAGTCAAGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1914
1861 GTGAGAGAAATGAGGAACTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 1920

QY	4081	GGCCGCGCATCCAGCCCGGAGAAAGAGCTTCACTCCAGAGAAATTAGTCGGGCTTT	414.4
Db	4135	GCCCCGCATCCAGCCCGGAGAAAGAGCTTCACTCCAGAGAAATTAGTCGGGCTTT	419.9
QY	4141	AAGGAAAGCATGCAACCAATATTTCTCAACGATTCAAAGTGAACAATGAGAGCC	420.0
Db	4195	AAGGAAAGCATGCAACCAATATTTCTCAACGATTCAAAGTGAACAATGAGAGCC	425.4
QY	4201	ACAAAGTGTGCTGTGTCTGGAATPCCGTGACCTTTGACAGCCAGAGCATCCAAATGTCTC	426.6
Db	4255	ACAAAGTGTGCTGTGTCTGGAATPCCGTGACCTTTGACAGCCAGAGCATCCAAATGTCTC	431.4
QY	4261	GAATGTCAAGTGTGTGTCAACCCCAAGTCTTCAAGTGTCTTCCAGGCCACTTCGGCTTG	432.0
Db	4315	GAATGTCAAGTGTGTGTGTCAACCCCAAGTCTTCAAGTGTCTTCCAGGCCACTTCGGCTTG	437.4
QY	4321	CCTGTGTAATPAGCCACACCTTCAACGAGGCTTTCTGSCGATGACAAATATGACCTCCCA	438.0
Db	4375	CCTGTGTAATPAGCCACACCTTCAACGAGGCTTTCTGSCGATGACAAATATGACCTCCCA	443.4
QY	4381	GCTCTCCAGACCAAGAGCCCAAGCAGACGCTTGCACTTGAAAGGCTGATGAAGTGTCC	444.0
Db	4435	GCTCTCCAGACCAAGAGCCCAAGCAGACGCTTGCACTTGAAAGGCTGATGAAGTGTCC	449.4
QY	4441	AGGAATTAACAACGAGGACAGCAAGGCTGGGACAGAGAGTACATGTCTCTGAGAGATCA	450.0
Db	4495	AGGAATTAACAACGAGGACAGCAAGGCTGGGACAGAGAGTACATGTCTCTGAGAGATCA	455.4
QY	4501	AAAGTCTCATTTATGACAAATGAGCCAGAGAGCTGGAACAGAGCCGCTGGAAGATTT	456.0
Db	4555	AAAGTCTCATTTATGACAAATGAGCCAGAGAGCTGGAACAGAGCCGCTGGAAGATTT	461.4
QY	4561	GAGCTGTGCTCTCCGACGGGGATGTATCTATTCAATGATGCGGTGGTCTTCGAACTC	462.0
Db	4615	GAGCTGTGCTCTCCGACGGGGATGTATCTATTCAATGATGCGGTGGTCTTCGAACTC	467.4
QY	4621	GCAATATCAGCCCAAGAGCATGTCTCCATACATCTGAAAGATGGAATCTCACCCGCAACC	468.0
Db	4675	GCAATATCAGCCCAAGAGCATGTCTCCATACATCTGAAAGATGGAATCTCACCCGCAACC	473.4
QY	4681	ACCTGTGAGCCCGGGAGAAACCTCTACTGTGATGTCCAGGTTCCTGACAAACAGAGC	474.0
Db	4735	ACCTGTGTGCTCCGGAGAAACCTCTACTGTGATGTCCAGGTTCCTGACAAACAGAGC	479.4
QY	4741	TGGGTCAACCGCTTGAATCAAGTGTCCAGAGTGGGAGAGTTTCTAGGGAATAAGCAGAA	480.0
Db	4795	TGGGTCAACCGCTTGAATCAAGTGTCCAGAGTGGGAGAGTTTCTAGGGAATAAGCAGAA	485.4
QY	4801	GCTGATGTCTAAACTCTCTTGAAAACTCCCTGCTGAACTGGAAGGATGAGACGCTTACAC	486.6
Db	4855	GCTGATGTCTAAACTCTCTTGAAAACTCCCTGCTGAACTGGAAGGATGAGACGCTTACAC	491.4
QY	4861	ATGAACTGCAAGCTGTCCCTTCAATGACAGAGTGTGTTGTGTGGGCAACCGAGAAAGGCTC	492.0
Db	4915	ATGAACTGCAAGCTGTCCCTTCAATGACAGAGTGTGTTGTGTGGGCAACCGAGAAAGGCTC	497.4
QY	4921	TACGCGCTGAATGTCTTGAAAAAACTCCCTAAACCATGTCCCGAGAAATTGGAGCAGTCTTC	498.0
Db	4975	TACGCGCTGAATGTCTTGAAAAAACTCCCTAAACCATGTCCCGAGAAATTGGAGCAGTCTTC	503.4
QY	4981	CAAAATTTATTTATACAGAGACTTGAAGAACTTACATGATAGCAGAGAAAGAGCGGAGCA	504.0
Db	5035	CAAAATTTATTTATACAGAGACTTGAAGAACTTACATGATAGCAGAGAAAGAGCGGAGCA	509.4
QY	5041	CTGTGTCTTGAGAGTGAAGAAAGTGAAGAGTCCCTGTAGCCAGTCCCACTGTGCTCC	510.0
Db	5095	CTGTGTCTTGAGAGTGAAGAAAGTGAAGAGTCCCTGTAGCCAGTCCCACTGTGCTCC	515.4
QY	5101	CAGCCCGACATCTCAACCCCAATTTTBAAGTGTCAAGAGGCTGTCACTTGTTTGGGACA	516.0
Db	5155	CAGCCCGACATCTCAACCCCAATTTTBAAGTGTCAAGAGGCTGTCACTTGTTTGGGACA	521.4

QY	5161	GGCAAGATTGAAGACGGGACTCTGATCATCTGAGACCCATATGCCAGCAAGTCATCTTC	5222
Db	5215	GGCAAGATTGAAGACGGGACTCTGATCATCTGAGACCATGCCAGCAAGTCATCTTC	5274
QY	5221	CGCTACACGAAACCTTAGCAATCTGCATCCGGAAGAAGATGAGACCTTAGACCC	5280
Db	5275	CGCTACACGAAACCTTAGCAATCTGCATCCGGAAGAAGATGAGACCTTAGACCC	5334
QY	5281	TGAGCGTGTATCCACTTACCAATTAAGATATCCTCATTTGGAAACCAATTAATTTCGAA	5340
Db	5335	TGAGCGTGTATCCACTTACCAATTAAGATATCCTCATTTGGAAACCAATTAATTTCGAA	5394
QY	5341	ATCGACATGAAAGCAGTACACGCTTGAGAAATTCCTGTGATTAAGATGACCAATTCCTGGCA	5400
Db	5395	ATCGACATGAAAGCAGTACACGCTTGAGAAATTCCTGTGATTAAGATGACCAATTCCTGGCA	5454
QY	5401	CCCTGTGTGTTTGCCGCTCTTCCACAAGCTTCCCTCTTCATACGTGAGGTGAAACAGC	5460
Db	5455	CCCTGTGTGTTTGCCGCTCTTCCACAAGCTTCCCTCTTCATACGTGAGGTGAAACAGC	5514
QY	5461	GCAGGGCAGCGAGAGAGTACTTGCCTGTGTGTTCCACGAATTTGGAGTTCGTGATTTCT	5520
Db	5515	GCAGGGCAGCGAGAGAGTACTTGCCTGTGTGTTCCACGAATTTGGAGTTCGTGATTTCT	5574
QY	5521	TACGGAAGCGTACCGCGACAGAGATCTAAGTGGAGTGGCTTACCTTTGACCTTTGCC	5580
Db	5575	TACGGAAGCGTACCGCGACAGAGATCTAAGTGGAGTGGCTTACCTTTGACCTTTGCC	5634
QY	5581	TACAGAGAACCCTATCTGTTTGTGACCCACTTCACTCACTCGAAGTATTTGAGATCCAG	5640
Db	5635	TACAGAGAACCCTATCTGTTTGTGACCCACTTCACTCACTCGAAGTATTTGAGATCCAG	5694
QY	5641	GCAGGCTCTCAGAGAGGAGACCCCGCCGAGCGTACTGAGATCCCGAACCAGCGCTAC	5700
Db	5695	GCAGGCTCTCAGAGAGGAGACCCCGCCGAGCGTACTGAGATCCCGAACCAGCGCTAC	5754
QY	5701	CTGGGCCCCCTGCATTTTCTCTCAGAGGCATTTACTTGGCGTCTCTCATACAGATTAATTA	5760
Db	5755	CTGGGCCCCCTGCATTTTCTCTCAGAGGCATTTACTTGGCGTCTCTCATACAGATTAATTA	5814
QY	5761	AGGTCATTTGCTCAGCAAGGAAACCTCGTGAAGATCCGGACCTGAACACACCGGGGC	5820
Db	5815	AGGTCATTTGCTCAGCAAGGAAACCTCGTGAAGATCCGGACCTGAACACACCGGGGC	5874
QY	5821	CCGTCCACTTCCCGCAGACGCCCCCAAGCGAGGCCACCAAGTACCAAGCAGACATC	5880
Db	5875	CCGTCCACTTCCCGCAGACGCCCCCAAGCGAGGCCACCAAGTACCAAGCAGACATC	5934
QY	5881	ACCAAGCGCGTGGCTCCAGACCCAGCGCGCCCGGAAGGCCACCCGCGAGAACCA	5940
Db	5935	ACCAAGCGCGTGGCTCCAGACCCAGCGCGCCCGGAAGGCCACCCGCGAGAACCA	5994
QY	5941	AGCACACCCCAACCCCTACCGGAGGGCGGACCGAGCTGGGACGAGGACAAATCTCTGGC	6000
Db	5995	AGCACACCCCAACCCCTACCGGAGGGCGGACCGAGCTGGGACGAGGACAAATCTCTGGC	6054
QY	6001	CGCCCCCTTGGAGCGAGAAATCCCCCGCGCGCGATATCTCAGACATCCGAGAGAGCGCTCC	6060
Db	6055	CGCCCCCTTGGAGCGAGAAATCCCCCGCGCGCGATATCTCAGACATCCGAGAGAGCGCTCC	6114
QY	6061	CCCGCGAGGCTGTTTGAAGACAGCAGCAGGGGCGGCTGGCTGGCGGAGCGCGTAGACCC	6120
Db	6115	CCCGCGAGGCTGTTTGAAGACAGCAGCAGGGGCGGCTGGCTGGCGGAGCGCGTAGACCC	6174
QY	6121	CCCGTGTCCCAAGGTGAACAAGTCTGGGACCAAGTCTTCAATATTA 6165	
Db	6175	CCCGTGTCCCAAGGTGAACAAGTCTGGGACCAAGTCTTCAATATTA 6219	

RESULT 3			
AX671105			
LOCUS	AX671105	6165 bp	DNA
DEFINITION	Sequence 1 from Patent WO03004523.	1linear	PAT 27-MAR-2003

QY 4021 ATGCCATGTCGGCCATGTCGGTGGCGAGACCAAGCCAGTGCATGAGCTGCTG 4080
 Db 4021 ATGCCATGTCGGCCATGTCGGTGGCGAGACCAAGCCAGTGCATGAGCTGCTG 4080
 QY 4081 GCCCGCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTT 4140
 Db 4081 GCCCGCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTT 4140
 QY 4141 AAGGAAAGCATGACCAATATTTCTCAACCATTTAAAGTGAAGTGAAGTGAAGCC 4200
 Db 4141 AAGGAAAGCATGACCAATATTTCTCAACCATTTAAAGTGAAGTGAAGTGAAGCC 4200
 QY 4201 ACAAGTGTCTGTGTCTGTGATCCGTCATCTTGACGTCGTCGTCGTCGTCGTC 4260
 Db 4201 ACAAGTGTCTGTGTGTCTGTGATCCGTCATCTTGACGTCGTCGTCGTCGTCGTC 4260
 QY 4261 GAATGTGAGTGTGTGTCACTCCCAAGTGTCTCACTGCTGCTGCTGCTGCTGCTG 4320
 Db 4261 GAATGTGAGTGTGTGTCACTCCCAAGTGTCTCACTGCTGCTGCTGCTGCTGCTG 4320
 QY 4321 CCGTGAATATATGACCACTTCACTGAGGCTTGTGCGTGAAGAAATGAAGTCCCA 4380
 Db 4321 CCGTGAATATATGACCACTTCACTGAGGCTTGTGCGTGAAGAAATGAAGTCCCA 4380
 QY 4381 GATCTCCAGACCAAGGAGCCAGACAGCTTGCACCTGGAAGATGAAGTGAAGTCCC 4440
 Db 4381 GATCTCCAGACCAAGGAGCCAGACAGCTTGCACCTGGAAGATGAAGTGAAGTCCC 4440
 QY 4441 AGGAATATCAAGAGAGACCAAGGCTGAGACAGAGATGACATTTCTCTGAGAGATCA 4500
 Db 4441 AGGAATATCAAGAGAGACCAAGGCTGAGACAGAGATGACATTTCTCTGAGAGATCA 4500
 QY 4501 AAGTCTCATTTATGACCAATGAAGCAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 4560
 Db 4501 AAGTCTCATTTATGACCAATGAAGCAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 4560
 QY 4561 GAGTGTGCTTCCCGAGGAGATGATATCTATCTATGATGATGATGATGATGATGATG 4620
 Db 4561 GAGTGTGCTTCCCGAGGAGATGATATCTATCTATGATGATGATGATGATGATGATG 4620
 QY 4621 GCAAAATACAGCAAGAGAGATGTCCTCACTACTGAAGTGAAGTGAAGTGAAGTGA 4680
 Db 4621 GCAAAATACAGCAAGAGAGATGTCCTCACTACTGAAGTGAAGTGAAGTGAAGTGA 4680
 QY 4681 ACCGTGTGCTCCCGAGAGACCTCTCTACTGATGATGATGATGATGATGATGATG 4740
 Db 4681 ACCGTGTGCTCCCGAGAGACCTCTCTACTGATGATGATGATGATGATGATGATG 4740
 QY 4741 TGGGTCAAGCCTTAGAATCAGTTGTGAGAGTGGAGAGTTCCTAGAGAGAGAGAGAA 4800
 Db 4741 TGGGTCAAGCCTTAGAATCAGTTGTGAGAGTGGAGAGTTCCTAGAGAGAGAGAA 4800
 QY 4801 GCTGTATCTAACTGCTTGAAGAACTCTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 4860
 Db 4801 GCTGTATCTAACTGCTTGAAGAACTCTCTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 4860
 QY 4861 ATGAACCTGACGCTGCTTCAAGTACCAAGTGTGTGTGAGAGAGAGAGAGAGAGTGC 4920
 Db 4861 ATGAACCTGACGCTGCTTCAAGTACCAAGTGTGTGTGAGAGAGAGAGAGAGAGTGC 4920
 QY 4921 TAGCCCTGATGTCTTGAAGAACTCCCTTAACCATATGTCAGAGAAATTTGAGCAGTCTT 4980
 Db 4921 TAGCCCTGATGTCTTGAAGAACTCCCTTAACCATATGTCAGAGAAATTTGAGCAGTCTT 4980
 QY 4981 CAAATTTATATTAATCAAGAGACTGAGAGAGTACTATGATACAGAGAGAGAGAGAGAG 5040
 Db 4981 CAAATTTATATTAATCAAGAGACTGAGAGAGTACTATGATACAGAGAGAGAGAGAGAG 5040
 QY 5041 CTGTGTCTTGTGAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5100
 Db 5041 CTGTGTCTTGTGAGAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5100
 QY 5101 CAGCCGACATCTACCAACCAATTTTGAAGTGTCAAGGCTGCCACTTGTGTGGAGCA 5160

Db 5101 CAGCCGACATCTACCAACCAATTTTGAAGTGTCAAGGCTGCCACTTGTGTGGAGCA 5160
 QY 5161 GGCAGAGATTGAAGAGGCTCTGTGATCTGTGACAGCAATGCCAGAGAAATGTCATTTCTC 5220
 Db 5161 GGCAGAGATTGAAGAGGCTCTGTGATCTGTGACAGCAATGCCAGAGAAATGTCATTTCTC 5220
 QY 5221 CGCTACAGAGAAATCTCAGAGAAATAGTGAATCCGAGAGAGAGAGAGAGAGAGAGAG 5280
 Db 5221 CGCTACAGAGAAATCTCAGAGAAATAGTGAATCCGAGAGAGAGAGAGAGAGAGAGAG 5280
 QY 5281 TGCAGCTGTATCTACCTTCAACCAATTAAGATCTTCAATTTGAGAAATTAATTTAGAA 5340
 Db 5281 TGCAGCTGTATCTACCTTCAACCAATTAAGATCTTCAATTTGAGAAATTAATTTAGAA 5340
 QY 5341 ATGCAATGAAGAGTACAGCTTGAAGATTTCTGATTAAGATTAAGATTAAGATTTAGCA 5400
 Db 5341 ATGCAATGAAGAGTACAGCTTGAAGATTTCTGATTAAGATTAAGATTAAGATTTAGCA 5400
 QY 5401 CCGTGTGTGTGAGAGCTTCTTCAAGAGCTTCCCTGTCTCAATCTGAGAGTGAAGAG 5460
 Db 5401 CCGTGTGTGTGAGAGCTTCTTCAAGAGCTTCCCTGTCTCAATCTGAGAGTGAAGAG 5460
 QY 5461 GAGAGGAGAGAGAGAGTACTGTGTGTGTTCAGAAATTTGAGAGTGTGTGTGTGTCT 5520
 Db 5461 GAGAGGAGAGAGAGAGTACTGTGTGTGTTCAGAAATTTGAGAGTGTGTGTGTGTCT 5520
 QY 5521 TAGGAGAGAGTACCGAGAGAGATCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5580
 Db 5521 TAGGAGAGAGTACCGAGAGAGATCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5580
 QY 5581 TACAGAGAACCTATCTGT 5640
 Db 5581 TACAGAGAACCTATCTGT 5640
 QY 5641 GAGAGCTCTGAG 5700
 Db 5641 GAGAGCTCTGAG 5700
 QY 5701 CTGAGGCTCTGAG 5760
 Db 5701 CTGAGGCTCTGAG 5760
 QY 5761 AAGGTCAATTTGCTGAG 5820
 Db 5761 AAGGTCAATTTGCTGAG 5820
 QY 5821 AGGTCATTTGCTGAG 5880
 Db 5821 AGGTCATTTGCTGAG 5880
 QY 5881 ACCAAGGCGGTGCTCTCAAGCAGCGCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 5940
 Db 5881 ACCAAGGCGGTGCTCTCAAGCAGCGCTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAG 5940
 QY 5941 AGCAACCCCAACCGCTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
 Db 5941 AGCAACCCCAACCGCTTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
 QY 6001 CGCCCTCTGAG 6060
 Db 6001 CGCCCTCTGAG 6060
 QY 6061 CCGGAG 6120
 Db 6061 CCGGAG 6120
 QY 6121 CCGTGTCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6185
 Db 6121 CCGTGTCTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6185
 RESULT 4

AX671108
LOCUS AX671108 8603 bp DNA linear PAT 27-MAR-2003
DEFINITION Sequence 4 from Patent WO03004523.
ACCESSION AX671108
VERSION AX671108.1 GI:29329570
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Zhu, Z.
TITLE Regulation of human cItron rho/rac-interacting kinase
JOURNAL Patent: WO 03004523-A 4 16-JAN-2003;
Bayer Aktienselschaft (DE)
FEATURES
Source location/Qualifiers
1..8603
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 2305 a 2206 c 2215 g 1877 t
ORIGIN
Query Match 99.9%; Score 6160.2; DB 6; Length 8603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTGGAAGTTCAATATGAGAGCGGAGATCTTTGGATGCTGCTGTAACCAT 60
Db 1 ATGTGGAAGTTCAATATGAGAGCGGAGATCTTTGGATGCTGCTGTAACCAT 60
QY 61 GCCAGCCGGGCTCCAGAGCTGATCTCTTCCAGGGGAAACACCCCTTATGACTCA 120
Db 61 GCCAGCCGGGCTCCAGAGCTGATCTCTTCCAGGGGAAACACCCCTTATGACTCA 120
QY 121 CAGCAGATGTCCTCTTCCCGAGAGGATATTAATGCCCTTGTCTCTTTGAA 180
Db 121 CAGCAGATGTCCTCTTCCCGAGAGGATATTAATGCCCTTGTCTCTTTGAA 180
QY 181 GAATGAGTCAAGCTCTCTGATGAGATTAAAGCAGTGAACCTTTGTCCGAAAT 240
Db 181 GAATGAGTCAAGCTCTCTGATGAGATTAAAGCAGTGAACCTTTGTCCGAAAT 240
QY 241 TCCGACACATAGCTGATGAGTCAAGAGCTCCAGGCTCCGCAAGAGCTTCCAGTCA 300
Db 241 TCCGACACATAGCTGATGAGTCAAGAGCTCCAGGCTCCGCAAGAGCTTCCAGTCA 300
QY 301 AGTCTTGAAGTTGTGTCATCTTTGCTGAAGTGAAGTGTGAAGAGCAACCGG 360
Db 301 AGTCTTGAAGTTGTGTCATCTTTGCTGAAGTGAAGTGTGAAGAGCAACCGG 360
QY 361 GACATCTAATGCTAATGAGTGAAGTGAAGAGGCTTATTTGGCCCGAGAGCTTTCA 420
Db 361 GACATCTAATGCTAATGAGTGAAGTGAAGAGGCTTATTTGGCCCGAGAGCTTTCA 420
QY 421 TTTTGTGAGAGAGCGGACATATTTATCTGAAGCAAGCCCGTGAATCCCCCAATTA 480
Db 421 TTTTGTGAGAGAGCGGACATATTTATCTGAAGCAAGCCCGTGAATCCCCCAATTA 480
QY 481 CAGTATGCTTTTCAAGCAAAAATCACCTTTATCTGTCATGGAATATCAGCTGAGG 540
Db 481 CAGTATGCTTTTCAAGCAAAAATCACCTTTATCTGTCATGGAATATCAGCTGAGG 540
QY 541 GACTTGTCTCATCTTTGATATGATGAGACCAAGTATGAAAACCTGATACGTTT 600
Db 541 GACTTGTCTCATCTTTGATATGATGAGACCAAGTATGAAAACCTGATACGTTT 600
QY 601 TACCTAGCTAGCTGATTTTGGCTTCAAGCGTTTCATGATGAGATACGTCAGCA 660
Db 601 TACCTAGCTAGCTGATTTTGGCTTCAAGCGTTTCATGATGAGATACGTCAGCA 660
QY 661 GACATCAAGCTGAGATTTCTGTTGACCGCAAGACACATCAAGCTGAGATTTT 720
Db 661 GACATCAAGCTGAGATTTCTGTTGACCGCAAGACACATCAAGCTGAGATTTT 720

Db 661 GACATCAAGCTGAGATTTCTGTTGACCGCAAGACACATCAAGCTGAGATTTT 720
QY 721 GATATCGCCGCAAAATGATTTCAACAGATGTTGAATGCCAACTCCGATTTGGAGC 780
Db 721 GATATCGCCGCAAAATGATTTCAACAGATGTTGAATGCCAACTCCGATTTGGAGC 780
QY 781 CCAGATTCATGAGCTCTGAAGTGTGATGTAACGGGGATGAAAAAGGACCTAC 840
Db 781 CCAGATTCATGAGCTCTGAAGTGTGATGTAACGGGGATGAAAAAGGACCTAC 840
QY 841 GGCCTGACCTGTGACTGTGTGCTGAGTGGCGCTGATTTGCTATGAGATATTTAGGAGA 900
Db 841 GGCCTGACCTGTGACTGTGTGCTGAGTGGCGCTGATTTGCTATGAGATATTTAGGAGA 900
QY 901 TCCCGCTTCGAGAGGAAACCTTCGCCGAACCTTCAATATATGAAATTTCCAGCG 960
Db 901 TCCCGCTTCGAGAGGAAACCTTCGCCGAACCTTCAATATATGAAATTTCCAGCG 960
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTCTTCTGATCTGATCAAGC 1020
Db 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTCTTCTGATCTGATCAAGC 1020
QY 1021 TTGTTGTGCGGCGAGAAAGAGAGTGAAGTTGAAGTCTTTGCTGCAATCTTTCTTC 1080
Db 1021 TTGTTGTGCGGCGAGAAAGAGAGTGAAGTTGAAGTCTTTGCTGCAATCTTTCTTC 1080
QY 1081 TCTAAATTTGACTGGAACAACATTTGTAATCTCTCCCTCCCTGTTCCCACTCAAG 1140
Db 1081 TCTAAATTTGACTGGAACAACATTTGTAATCTCTCTCCCTCCCTGTTCCCACTCAAG 1140
QY 1141 TCTGACGATGACCTCCAAATTTGATGAACCAAGAAATTCGTGGTTTCACTCT 1200
Db 1141 TCTGACGATGACCTCCAAATTTGATGAACCAAGAAATTCGTGGTTTCACTCT 1200
QY 1201 CCGTGCAGCTGAGAGCCCTCAGAGCTTCCGGGTGAAGAACTGGGCTTTG 1260
Db 1201 CCGTGCAGCTGAGAGCCCTCAGAGCTTCCGGGTGAAGAACTGGGCTTTG 1260
QY 1261 TACAGCAAGCACTGGGATTTCTTGATGATCTGATCTGTTGTGTCGGATCTGACCTC 1320
Db 1261 TACAGCAAGCACTGGGATTTCTTGATGATCTGATCTGTTGTGTCGGATCTGACCTC 1320
QY 1321 CCGTGCAGCTGAGCTCAATGAAAAAGAACTTCTCATCAAAAGCAAGAGCTACAAAC 1380
Db 1321 CCGTGCAGCTGAGCTCAATGAAAAAGAACTTCTCATCAAAAGCAAGAGCTACAAAC 1380
QY 1381 TCTCAGAGCAAGTGTCAAGAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
Db 1381 TCTCAGAGCAAGTGTCAAGAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
QY 1441 GAGGTGAGGCTGTGCTTATGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db 1441 GAGGTGAGGCTGTGCTTATGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
QY 1501 TCCCTCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Db 1501 TCCCTCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 1561 TTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
Db 1561 TTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
QY 1621 GATATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
Db 1621 GATATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
QY 1681 GTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Db 1681 GTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
QY 1741 CGGAGTGTCTCTGAGATCTGAGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
Db 1741 CGGAGTGTCTCTGAGATCTGAGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800


```
QY 1801 CGGAAAGCGACAGATGTGAGATTAACCTGTTGAAGCTAAGAGATCAAGGAAAGCTTGA 1860
DB 1801 CGGAAAGCGACAGAAATGTGAGATTAACCTGTTGAAGCTAAGAGATCAAGGAAAGCTTGA 1860
QY 1861 GTGGGGAATATGCGAAATCTGGAGAAATCAATGCTGAGCAGACCTCAAAATTCAGAG 1920
DB 1861 GTGGGGAATATGCGAAATCTGGAGAAATCAATGCTGAGCAGACCTCAAAATTCAGAG 1920
QY 1921 CTCGAAGAGAACTGGAGAAAGCTGTAAAGCCAGACCGAGGCCAGAGCTGCTGAG 1980
DB 1921 CTCGAAGAGAACTGGAGAAAGCTGTAAAGCCAGACCGAGGCCAGAGCTGCTGAG 1980
QY 1981 AATATCCCGCAGGCAAAAGAGCGAGCCGAGAGGAGCTGAGAAAGCTGCAAAACCGAG 2040
DB 1981 AATATCCCGCAGGCAAAAGAGCGAGCCGAGAGGAGCTGAGAAAGCTGCAAAACCGAG 2040
QY 2041 GATTTCTTGAAGGCTCAGAAAGAAAGCTGTGAGAAAGCTGAGAAAGCCGCAATCTCTG 2100
DB 2041 GATTTCTTGAAGGCTCAGAAAGAAAGCTGTGAGAAAGCTGAGAAAGCCGCAATCTCTG 2100
QY 2101 GAGAACAGGTAAAGAGCTAGAGACCATGAGCCGTAGAGAAACAGACTGAGAGTAGAC 2160
DB 2101 GAGAACAGGTAAAGAGCTAGAGACCATGAGCCGTAGAGAAACAGACTGAGAGTAGAC 2160
QY 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGAGCTGATTAATTTCTGAGCTCGAAGAG 2220
DB 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGAGCTGATTAATTTCTGAGCTCGAAGAG 2220
QY 2221 AAACATCGGAGGCGCCAAAGCTCTGAGCCAGCACTGAGAAAGCTGCAAAACAGAAAGAG 2280
DB 2221 AAACATCGGAGGCGCCAAAGCTCTGAGCCAGCACTGAGAAAGCTGCAAAACAGAAAGAG 2280
QY 2281 CAGCACTATGAGAGAAAGATTAAGTTAAGTTGAGCAATCAATTAAGAAAGAAAGCTGCTAG 2340
DB 2281 CAGCACTATGAGAGAAAGATTAAGTTAAGTTGAGCAATCAATTAAGAAAGAAAGCTGCTAG 2340
QY 2341 AAGAGACACTGAGAGAACTGATGAGAGACAGAGAGAGAGCCCATAGAGAGGCAAA 2400
DB 2341 AAGAGACACTGAGAGAACTGATGAGAGACAGAGAGAGAGCCCATAGAGAGGCAAA 2400
QY 2401 ATTCTAGGGAACAGAGAGCGATGATCAATGCTATGAGATTCAGAAATTCAGATCCCTGGA 2460
DB 2401 ATTCTAGGGAACAGAGAGCGATGATCAATGCTATGAGATTCAGAAATTCAGATCCCTGGA 2460
QY 2461 CAGAGATTGTGAACTGCTGAGAGCCAAATTAATTTGACAAATAGCAGTCTTTTACC 2520
DB 2461 CAGAGATTGTGAACTGCTGAGAGCCAAATTAATTTGACAAATAGCAGTCTTTTACC 2520
QY 2521 CAAAGAACATGAGAGCCCAAGAGAGATGATTTCTGAACTAGAGCAACAGAAATTTAC 2580
DB 2521 CAAAGAACATGAGAGCCCAAGAGAGATGATTTCTGAACTAGAGCAACAGAAATTTTAC 2580
QY 2581 CTGAGACAGAGGCTGGAAAGTTGAGAGCCCAAGAACCAAGAACTGAGAGAGAGCTGAG 2640
DB 2581 CTGAGACAGAGGCTGGAAAGTTGAGAGCCCAAGAACCAAGAACTGAGAGAGAGCTGAG 2640
QY 2641 AAGATAGCCACCAAGACCAACAGTGAACAGATCGCTGCTGAGAACTGAGAGCAAGATTG 2700
DB 2641 AAGATAGCCACCAAGACCAACAGTGAACAGATCGCTGCTGAGAACTGAGAGCAAGATTG 2700
QY 2701 CGGAGAGTCAATGAGACAGAGAGCAAGAACTGAGAGCTCAAGAGCCAGCTCAAGAG 2760
DB 2701 CGGAGAGTCAATGAGACAGAGAGCAAGAACTGAGAGCTCAAGAGCCAGCTCAAGAG 2760
QY 2761 CTACAGCTCTCCCTGAGAGAGCGAGTCAAGATTGACAGCCCTGAGAGCTGAGAGAG 2820
DB 2761 CTACAGCTCTCCCTGAGAGAGCGAGTCAAGATTGACAGCCCTGAGAGCTGAGAGAG 2820
QY 2821 GCGCTGAGAGAGCGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2880
DB 2821 GCGCTGAGAGAGCGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2880
```


RESULT 5
 AX671112 6156 bp DNA linear PAT 27-MAR-2003
 LOCUS
 DEFINITION Sequence 8 from Patent WO03004523.
 AX671112
 ACCSSION
 VERSION AX671112.1 GI:29329572
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 Zhu Z.
 Regulation of human citron rho/rac-interacting kinase
 TITLE Patent: WO 03004523-A 8 16-JAN-2003;
 JOURNAL Bayer Aktiengesellschaft (DE)
 FEATURES
 source 1..6156
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 1732 a 1548 c 1679 g 1197 t
 ORIGIN
 Query Match 99.6%; Score 6142.2; DB 6; Length 6156;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 6147; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 1 ATGTTGAAGTTCAATATGAGCGCGGAATCCTTTGGATGCTGTGCTGTAACCCATT 60
 DB 1 ATGTTGAAGTTCAATATGAGCGCGGAATCCTTTGGATGCTGTGCTGTAACCCATT 60
 QY 61 GCCAGCCGGGCGCTCCAGGCTGAATCTGTTCTTCCAGAGGAAACCAACCTTTATGACTCA 120
 DB 61 GCCAGCCGGGCGCTCCAGGCTGAATCTGTTCTTCCAGAGGAAACCAACCTTTATGACTCA 120
 QY 121 CAGCAGATGCTCTCTTCCGAGAGGATATAGATGCCCTCTTTGTTCTTTGAA 180
 DB 121 CAGCAGATGCTCTCTTCCGAGAGGATATAGATGCCCTCTTTGTTCTTTGAA 180
 QY 181 GAATGAGTCAGCTGCTGATGAGATTAAGCACTGAGCACTTTGTCGGAAGTAT 240
 DB 181 GAATGAGTCAGCTGCTGATGAGATTAAGCACTGAGCACTTTGTCGGAAGTAT 240
 QY 241 TCCGACACCATAGCTAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 300
 DB 241 TCCGACACCATAGCTAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 300
 QY 301 AGCTTGTAGTGTGTGCTCACTTTGCTGAAGTCAGAGTGTAAAGAGAAACCAACCGGG 360
 DB 301 AGCTTGTAGTGTGTGCTCACTTTGCTGAAGTCAGAGTGTAAAGAGAAACCAACCGGG 360
 QY 361 GACATCTATGCTATGAAAGTATGAGAGAAAGGCTTTATGAGCCGACAGAGAGTTTCA 420
 DB 361 GACATCTATGCTATGAAAGTATGAGAGAAAGGCTTTATGAGCCGACAGAGAGTTTCA 420
 QY 421 TTTTGGAGAGAGCGGAAATATATTCGAGGACAGCCCGTGGATCCGCCCAATTA 480
 DB 421 TTTTGGAGAGAGCGGAAATATATTCGAGGACAGCCCGTGGATCCGCCCAATTA 480
 QY 481 CAGTATGCTTTGAGAGCAAAATACCTTTATCTGTGATGAGAAATATCAGGCTGAGGG 540
 DB 481 CAGTATGCTTTGAGAGCAAAATACCTTTATCTGTGATGAGAAATATCAGGCTGAGGG 540
 QY 541 GACTTGTGCTCACTTTTGAATATATGAGACCAAGTTAGTAAACCTGATACATTT 600
 DB 541 GACTTGTGCTCACTTTTGAATATATGAGACCAAGTTAGTAAACCTGATACATTT 600
 QY 601 TACCTGAGTGAAGTATTTGGCTGTTCAAGGCTCATCTGATGGAGTAACGTGATCA 660
 DB 601 TACCTGAGTGAAGTATTTGGCTGTTCAAGGCTCATCTGATGGAGTAACGTGATCA 660

QY 661 GACATCAAGCTTGAGAAACATTTCTGTTGACCCGACAGAGACATCAAGCTGTGATTTT 720
 DB 661 GACATCAAGCTTGAGAAACATTTCTGTTGACCCGACAGAGACATCAAGCTGTGATTTT 720
 QY 721 GGATCTGCGGAAATGAAATTCAAACAAATGATGTAATCCAAACCTCCGATTTGGAGC 780
 DB 721 GGATCTGCGGAAATGAAATTCAAACAAATGATGTAATCCAAACCTCCGATTTGGAGC 780
 QY 781 CCAATATTAATGAGCTCTGTAAGTGTGACTGTGATGAGACGAGAGTGGAAAAAGCACCTAC 840
 DB 781 CCAATATTAATGAGCTCTGTAAGTGTGACTGTGATGAGACGAGAGTGGAAAAAGCACCTAC 840
 QY 841 GGCTTGACCTGTGACCTGTGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
 DB 841 GGCTTGACCTGTGACCTGTGATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
 QY 901 TCCCTCTTCCGAGAGGAACTCTGCGAGAACCTTCATTAATCAATTAATTTCCAGCGG 960
 DB 901 TCCCTCTTCCGAGAGGAACTCTGCGAGAACCTTCATTAATCAATTAATTTCCAGCGG 960
 QY 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAATTTTGAATTTTGAATTTTGAATTTTGA 1020
 DB 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAATTTTGAATTTTGAATTTTGAATTTTGA 1020
 QY 1021 TTTGTTGCGGCGCAGAAAGAGAGACTGAATTTGAAGTCTTGTGCTGCTCAATCCTTTCTTC 1080
 DB 1021 TTTGTTGCGGCGCAGAAAGAGAGACTGAATTTGAAGTCTTGTGCTGCTCAATCCTTTCTTC 1080
 QY 1081 TCTTAAATTTGACTGGAACAAATTTGTAATCTTCTTCCCTTGTGTTCCACCTTAAG 1140
 DB 1081 TCTTAAATTTGACTGGAACAAATTTGTAATCTTCTTCCCTTGTGTTCCACCTTAAG 1140
 QY 1141 TCTGACGATGACCTTCAATTTTGAATGACCAAGAAATTTGATGATTTGATGATTTGATGAT 1200
 DB 1141 TCTGACGATGACCTTCAATTTTGAATGACCAAGAAATTTGATGATTTGATGATTTGATGAT 1200
 QY 1201 CCGTGCCAGCTGAGCCCTCTGAGGCTTCTGAGGATTTGATGATTTGATGATTTGATGAT 1260
 DB 1201 CCGTGCCAGCTGAGCCCTCTGAGGCTTCTGAGGATTTGATGATTTGATGATTTGATGAT 1260
 QY 1261 TAGACGAAGGACCTGAGGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 DB 1261 TAGACGAAGGACCTGAGGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 QY 1321 CCTGCAAGACTAGCTCCATGAAAGAACTTCTCATCAAAAGCAAAAGAGCTTACAGAC 1380
 DB 1321 CCTGCAAGACTAGCTCCATGAAAGAACTTCTCATCAAAAGCAAAAGAGCTTACAGAC 1380
 QY 1381 TCTTGAAGCAAGTGTCAAGATGAGACAGAGAAATGATCCGGTTACATCGGAGAGTGTCA 1440
 DB 1381 TCTTGAAGCAAGTGTCAAGATGAGACAGAGAAATGATCCGGTTACATCGGAGAGTGTCA 1440
 QY 1441 GAGGTGAGAGGCTGTGTTAGTCAAGAGAGTGTGAGCTGAGAGCTTGTGAGCTCAGAGA 1500
 DB 1441 GAGGTGAGAGGCTGTGTTAGTCAAGAGAGTGTGAGCTGAGAGCTTGTGAGCTCAGAGA 1500
 QY 1501 TCCCTCTGAGAGGAGCTTTGCTTACTATCATACAGATGCAAGTGTGAAAGCTTAAAGCAAGT 1560
 DB 1501 TCCCTCTGAGAGGAGCTTTGCTTACTATCATACAGATGCAAGTGTGAAAGCTTAAAGCAAGT 1560
 QY 1561 TTTGAGCAGAGCAGATGAGAGTGTGCTCCAGAGAGATGACAAAGCACTGAGCTTCTTCAT 1620
 DB 1561 TTTGAGCAGAGCAGATGAGAGTGTGCTCCAGAGAGATGACAAAGCACTGAGCTTCTTCAT 1620
 QY 1621 GATATCAGAGAGAGCTGAGAGCTTCCAGAAATCAAGAGAGAGAGTACAGAGCTCAA 1680
 DB 1621 GATATCAGAGAGAGCTGAGAGCTTCCAGAAATCAAGAGAGAGAGTACAGAGCTCAA 1680
 QY 1681 GTGAGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 DB 1681 GTGAGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

1741 CGGATGATCTTACGAAATCTGAGAGAGCTCGGCTTGCTGTAAGAATTCAAG 1800
1741 CGGATGATCTTACGAAATCTGAGAGAGCTCGGCTTGCTGTAAGAATTCAAG 1800
1801 CGGAAAGCGA CAGAAATCTGACATTAATCTGTAAGGCTTAAGATCAAGGAAAGCTGAA 1860
1801 CGGAAAGCGA CAGAAATCTGACATTAATCTGTAAGGCTTAAGATCAAGGAAAGCTGAA 1860
1861 GTGGAGAAATATGGGAAATCTGAGAGAGATCAATGCTGAGAGAGCTCAAAATTCAGAG 1920
1861 GTGGAGAAATATGGGAAATCTGAGAGAGATCAATGCTGAGAGAGCTCAAAATTCAGAG 1920
1921 CTCGAAAGAAATCTGAGAGAGCTGTAAGAAAGCGAGAGCGAGAGCTGCTGAG 1980
1921 CTCGAAAGAAATCTGAGAGAGCTGTAAGAAAGCGAGAGCGAGAGCTGCTGAG 1980
1981 AATATCCGCGAGAAAGAGAGCGAGAGCGAGAGCTGTAAGAAAGCGAGAG 2040
1981 AATATCCGCGAGAAAGAGAGCGAGAGCGAGAGCTGTAAGAAAGCGAGAG 2040
2041 GATTCCTTCTGAGAGAGCTGTAAGAAAGCGAGAGCGAGAGCTGTAAGAG 2100
2041 GATTCCTTCTGAGAGAGCTGTAAGAAAGCGAGAGCGAGAGCTGTAAGAG 2100
2101 GAGAAAGAGTAAAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAG 2160
2101 GAGAAAGAGTAAAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAG 2160
2161 ATCCAGACAAATATCCAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAG 2220
2161 ATCCAGACAAATATCCAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAG 2220
2221 AAACATCGGAG 2280
2221 AAACATCGGAG 2280
2281 CAGACATATGAG 2340
2281 CAGACATATGAG 2340
2341 AAG 2400
2341 AAG 2400
2401 ATTCTGAG 2460
2401 ATTCTGAG 2460
2461 CAG 2520
2461 CAG 2520
2521 CAAAG 2580
2521 CAAAG 2580
2581 CTGAG 2640
2581 CTGAG 2640
2641 AAGATCAG 2700
2641 AAGATCAG 2700
2701 CGGAG 2760
2701 CGGAG 2760
2761 CTAAG 2820
2761 CTAAG 2820
2821 GCCCTGAG 2880
2821 GCCCTGAG 2880

2821 GCCCTGAG 2880
2881 GAG 2940
2881 GAG 2940
2941 CTTGTAAG 3000
2941 CTTGTAAG 3000
3001 GACAG 3060
3001 GACAG 3060
3061 GAG 3120
3061 GAG 3120
3121 ACGAG 3180
3121 ACGAG 3180
3181 TGCAAG 3240
3181 TGCAAG 3240
3241 AAAG 3300
3241 AAAG 3300
3301 TGTGAG 3360
3301 TGTGAG 3360
3361 GATCAG 3420
3361 GATCAG 3420
3421 GAGATTCGAG 3480
3421 GAGATTCGAG 3480
3481 GACAG 3540
3481 GACAG 3540
3541 CAG 3600
3541 CAG 3600
3601 TTAAG 3660
3601 TTAAG 3660
3661 GAG 3720
3661 GAG 3720
3721 GAAAG 3780
3721 GAAAG 3780
3781 CAAAG 3840
3781 CAAAG 3840
3841 GTTCTGAG 3900
3841 GTTCTGAG 3900
3901 GAGCTAG 3960
3901 GAGCTAG 3960

QY 6121 CCGCTGTCCAGGTGAACAGTCTGTGGACCACTC 6155
Db 6121 CCGCTGTCCAGGTGAACAGTGAAGCGACGATTC 6155

RESULT 6
AX166510 6159 bp DNA linear PAT 22-JUN-2001
LOCUS Sequence 1 from Patent WO0138503.
DEFINITION AX166510
ACCESSION AX166510
VERSION AX166510.1 GI:14546855
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Plowman, G.D., Whyte, D., Manning, G.S., Sudarsanam, S.S., Martinez, R.,
Plasman, P. and Clary, D.S.
Novel human protein kinases and protein kinase-like enzymes
Patent: WO 0138503-A 1 31-MAY-2001.
Sugen, Inc. (US)
FEATURES
Location/Qualifiers
1..6159
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1732 a 1549 c 1680 g 1198 t
ORIGIN

Query Match 99.4%; Score 6126; DB 6; Length 6159;
Best Local Similarity 99.8%; Freq. No. 0;
Matches 6145; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

QY 1 ATGTGAAGTCAATATATGAGGCGCGAAATCCCTTGTGATGCTGTGCTCAACCAT 60
Db 1 ATGTGAAGTCAATATGAGGCGCGAAATCCCTTGTGATGCTGTGCTCAACCAT 60

QY 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTTCCAGGGGAAACACCTTTATGACTCA 120
Db 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTTCCAGGGGAAACACCTTTATGACTCA 120

QY 121 CACGAGTGTCTCTCTTCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180
Db 121 CACGAGTGTCTCTCTTCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180

QY 121 CACGAGTGTCTCTCTTCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180
Db 121 CACGAGTGTCTCTCTTCCGAGAGGATATTAGATGCCCTCTTGTCTCTTGA 180

QY 181 GAATGCAAGTCAAGCTGTGATGAAGATTAAACAGTGAACAATTGTCCGAA--G 237
Db 181 GAATGCAAGTCAAGCTGTGATGAAGATTAAACAGTGAACAATTGTCCGAAAGT 240

QY 238 TATTCGACACCATAGCTGATGATACAGAGCTCCAGCTTCGCAAGGACTTCGAGTC 297
Db 241 TATTCGACACCATAGCTGATGATACAGAGCTCCAGCTTCGCAAGGACTTCGAGTC 300

QY 298 AGAAGCTTGTAGATGTGTGCTCACTTGTGAAAGTCAAGTGTGATGAGAGAAAGCA 357
Db 301 AGAAGCTTGTAGATGTGTGCTCACTTGTGAAAGTCAAGTGTGATGAGAGAAAGCA 360

QY 358 GGGGACATCTATGCTTGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417
Db 361 GGGGACATCTATGCTTGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 418 TCAATTTTGGAG 477
Db 421 TCAATTTTGGAG 480

QY 478 TTACAGTATGCTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
Db 481 TTACAGTATGCTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540

QY 538 GGGGACCTGCTGCTCACTTTGAATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
Db 541 GGGGACCTGCTGCTCACTTTGAATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

QY 598 TTTTACCTAGCTAGCTGATTTTGGCTGTCAAGCTGATCTGATGGGATACGTCAT 657
Db 601 TTTTACCTAGCTAGCTGATTTTGGCTGTCAAGCTGATCTGATGGGATACGTCAT 660

QY 658 CGAGACATCAAGCTTGAGAGACATTTCTGTTACCGCAGAGACATCAAGCTGTGAT 717
Db 661 CGAGACATCAAGCTTGAGAGACATTTCTGTTACCGCAGAGACATCAAGCTGTGAT 720

QY 718 TTTGATCTGCTGGGAG 777
Db 721 TTTGATCTGCTGGGAG 780

QY 778 ACCCAGATTCATAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
Db 781 ACCCAGATTCATAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840

QY 838 TAGGCTGCTGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
Db 841 TAGGCTGCTGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

QY 898 AGATCCCTGCTGAG 957
Db 901 AGATCCCTGCTGAG 960

QY 958 CGGTTTTTGAATTTCCAGATGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
Db 961 CGGTTTTTGAATTTCCAGATGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

QY 1018 AGCTTTGTTGGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
Db 1021 AGCTTTGTTGGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

QY 1078 TTCTCTAATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137
Db 1081 TTCTCTAATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

QY 1138 AAGTCTGAGATGACCTTCAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
Db 1141 AAGTCTGAGATGACCTTCAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

QY 1198 TCTCGGAGCTGAG 1257
Db 1201 TCTCGGAGCTGAG 1260

QY 1258 TCGTACAG 1317
Db 1261 TCGTACAG 1320

QY 1318 TCCCTGCGAAG 1377
Db 1321 TCCCTGCGAAG 1380

QY 1378 GACTCTGAG 1437
Db 1381 GACTCTGAG 1440

QY 1438 TCAAGAGTGAAGCTGTGCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1441 TCAAGAGTGAAGCTGTGCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

QY 1498 AGATCCCTGCTGAG 1557
Db 1501 AGATCCCTGCTGAG 1560

QY 1558 AGTTTGAAG 1617
Db 1561 AGTTTGAAG 1620

QY 1618 CATGATATCAG 1677
Db 1621 CATGATATCAG 1680

Db	3841	AAAGTTCTCTGCACTACAAATGAGCTGAAGCTGGCCCTCGAGAAAGAAAGAAAGCTCGCTG	3900
Qy	3898	GCAGAGCTAGAGGAAGCCCTTCAAGAGACCCGCATCGAGCTCCGATCCGCCCGAGAGAA	3957
Db	3901	GCAGAGCTAGAGGAAGCCCTTCAAGAGACCCGCATCGAGCTCCGATCCGCCCGAGAGAA	3960
Qy	3998	GCTGCCACCGCAAMAGCAACGACCAACCCATCTACATGCCAGCCACCCGCAAGCAG	4017
Db	3961	GCTGCCACCGCAAMAGCAACGACCAACCCATCTACAGCCAGCCACCGCAAGCAG	4020
Qy	4018	CAGATCCGCATGTCCGCATCTGTGGGTGCGCAGAGCACAGACCAGGAGCCATGAGCTG	4077
Db	4021	CAGATCCGCATGTCCGCATCTGTGGGTGCGCAGAGCACAGACCAGGAGCCATGAGCTG	4080
Qy	4078	CTGGCCCCCGCATCCAGCGCGCAAGAAAGAGTCTTCACTCCAGAGAAATTATGTCGCGT	4137
Db	4081	CTGGCCCCCGCATCCAGCGCGCAAGAAAGAGTCTTCACTCCAGAGAAATTATGTCGCGT	4140
Qy	4138	CTTAAGAGAGCGATGCAACACAAATATCTCTACCGATTCAACGTATAGACTGAAATGCA	4197
Db	4141	CTTAAGAGAGCGATGCAACCAATATCTCTACCGATTCAACGTATAGACTGAAATGCA	4200
Qy	4198	GCACAAAGTGTGCTGTGTGTCTGATACCGTGCATTGGACCCAGCATCCAAATGT	4257
Db	4201	GCACAAAGTGTGCTGTGTGTCTGATACCGTGCATTGGACCGCAAGCATCCAAATGT	4260
Qy	4258	CTCGAATGTCAGGTGATGTGTCAACCCCAAGTGTCCACGTCCTGCCAGCCACTGTGGC	4317
Db	4261	CTCGAATGTCAGGTGATGTGTCAACCCCAAGTGTCCACGTCCTGCCAGCCACTGTGGC	4320
Qy	4318	TTGCTGTGTGATATATGCACACACTTCCAGAGGCTTGTCCGCTGACAAATGACATCC	4377
Db	4321	TTGCTGTGTGATATATGCACACACTTCCAGAGGCTTGTCCGCTGACAAATGACATCC	4380
Qy	4378	CCAGGCTTCCAGACCAAGAGCCCGACGACGCTTGCACCTGGAAAGGTGATGAAGTG	4437
Db	4381	CCAGGCTTCCAGACCAAGAGCCCGACGACGCTTGCACCTGGAAAGGTGATGAAGTG	4440
Qy	4438	CCCAAGATTAACAAACGAGGACAGCAAGAGGTGGGACAGAAAGTCAATTGTCTTGAAGGA	4497
Db	4441	CCCAAGATTAACAAACGAGGACAGCAAGAGGTGGGACAGAAAGTCAATTGTCTTGAAGGA	4500
Qy	4498	TCAAAAGTCTCATTTATGACATGAAACCAAGAGCTGGACAGAGCCGGTGAAGAA	4557
Db	4501	TCAAAAGTCTCATTTATGACATGAAACCAAGAGAGCTGGACAGAGCCGGTGAAGAA	4560
Qy	4558	TTTGAAGCTGTGCTTCCCGACGCGGAGTGTATCTTCAATGATGCGCTGTGGTCTCCGA	4617
Db	4561	TTTGAAGCTGTGCTTCCCGACGCGGAGTGTATCTTCAATGATGCGCTGTGGTCTCCGA	4620
Qy	4618	CTCGCAATACAGCCAAAGACAGATGTCCCATACATACCTGAAGATGGAATCACAACCCAC	4677
Db	4621	CTCGCAATACAGCCAAAGACAGATGTCCCATACATACCTGAAGATGGAATCACAACCCAC	4680
Qy	4678	AACCACTGTGTGGCCCGGAGAAACCTCTACTTGTCTAGTCCCACTTCCCTGACAAACG	4737
Db	4681	AACCACTGTGTGGCCCGGAGAAACCTCTACTTGTCTAGTCCCACTTCCCTGACAAACG	4740
Qy	4738	CGCTGGGTACACGCGCTTGAATTCAGATGTGTCCAGATGGGAGATTTCTAGGAGAAAGCA	4797
Db	4741	CGCTGGGTACACGCGCTTGAATTCAGATGTGTGTCCAGATGGGAGATTTCTAGGAGAAAGCA	4800
Qy	4798	GAACTGTATGCTAAACTGTCTTGAAACTCTCCTGCTGAACTGSAAGGTATGACCGTCTA	4857
Db	4801	GAACTGTATGCTAAACTGTCTTGAAACTCTCCTGCTGAACTGSAAGGTATGACCGTCTA	4860
Qy	4858	GACATGAACTCAACGCTGACCTTCAAGTATCACAGGTGGTGTGGTGGGACCCGAGAAAGG	4917
Db	4861	GACATGAACTCAACGCTGACCTTCAAGTATCACAGGTGGTGTGGTGGGACCCGAGAAAGG	4920
Qy	4918	CTCTAACGCTGAAATGTCTTGAATACTCCTTAACCAATGTCCAGAAATTGAGACGTG	4977
Db	4921	CTCTAACGCTGAAATGTCTTGAATACTCCTTAACCAATGTCCAGAAATTGAGACGTG	4980

QY	4978	TTCCAAATTTATATTAATCAAGACCTTGAGAGCTTACTCATGATAGACGAAAGACCG	5037
Db	4981	TTCCAAATTTATATTAATCAAGACCTTGAGAGCTTACTCATGATAGACGAAAGACCG	5040
QY	5038	GACCTGTCTCTGTGGACGTGAAGAAAGTGAACAGTCCCTGGCCCACTCCACTCTGCT	5097
Db	5041	GACCTGTCTCTGTGGACGTGAAGAAAGTGAACAGTCCCTGGCCCACTCCACTCTGCT	5100
QY	5098	GCCCAAGCCGACATCTCAACCAATTTTGAAGCTGTCAAGGGCTGCGCACTTGTGGG	5157
Db	5101	GCCCAAGCCGACATCTCAACCAATTTTGAAGCTGTCAAGGGCTGCGCACTTGTGGG	5160
QY	5158	GAGGCAAGATTGGAACGGGCTGGCATCTGTGCAGCATGGCCAGCAAAAGTGTCATT	5217
Db	5161	GAGGCAAGATTGGAACGGGCTGGCATCTGTGTGAGCATGGCCAGCAAAAGTGTCATT	5220
QY	5218	CTCCGCTACAAAGAAAACCTTGACAAATACTGCATCCGAAAAAGATAGAGACCTCGAG	5277
Db	5221	CTCCGCTACAAAGAAAACCTTGACAAATACTGCATCCGAAAAAGATAGAGACCTCGAG	5280
QY	5278	CCCTGACAGCTGTATCCACTTCAACCAATTAAGTATCCCTCATTTGAGCCAAATTAATTCAC	5337
Db	5281	CCCTGACAGCTGTATCCACTTCAACCAATTAAGTATCCCTCATTTGAGCCAAATTAATTCAC	5340
QY	5338	GAATTCGACATGAAGCAGTACACGCTGAGAGAAATTCCTGATTAAGAAATGACATTCTTG	5397
Db	5341	GAATTCGACATGAAGCAGTACACGCTGAGAGAAATTCCTGATTAAGAAATGACATTCTTG	5400
QY	5398	GCACCTGTGTGTGTGGCCGCTCTTCCAAACAGTCTCCCTGTCCATATCGACAGGTGAAC	5457
Db	5401	GCACCTGTGTGTGTGGCCGCTCTTCCAAACAGTCTCCCTGTCCATATCGTGCAGGTGAAC	5460
QY	5458	AGCCGAGGGCAGCGAGAGAGAGTACTTGCTGTGTTTCCACGAAATTTGAGATGTTCGTGAT	5517
Db	5461	AGCCGAGGGCAGCGAGAGAGAGTACTTGCTGTGTTTCCACGAAATTTGAGATGTTCGTGAT	5520
QY	5518	TCTTAACGAAACGTAGCCGACGACAGAGATCTCAAGTGGAGTCCCTTAACCTTGGACCTTT	5577
Db	5521	TCTTAACGAAACGTAGCCGACGACAGAGATCTCAAGTGGAGTCCCTTAACCTTGGACCTTT	5580
QY	5578	GCCTACAGAGAACCTTATCTGTGTTGTGACCCACTTCAACTCTCATCTGAGATTAATGAGATC	5637
Db	5581	GCCTACAGAGAACCTTATCTGTGTTGTGACCCACTTCAACTCTCATCTGAGATTAATGAGATC	5640
QY	5638	CAGGCAACGCTCCTCAGACAGGAGACCCCTGACCAGACGTACTGGACATCCCGAACCCGCGC	5697
Db	5641	CAGGCAACGCTCCTCAGACAGGAGACCCCTGACCAGACGTACTGGACATCCCGAACCCGCGC	5700
QY	5698	TACTGTGGGCTCTGCATTTCTCTCAAGAGCGATTTACTTGGGCTCTCATACAGGATTA	5757
Db	5701	TACTGTGGGCTCTGCATTTCTCTCAAGAGCGATTTACTTGGGCTCTCATACAGGATTA	5760
QY	5758	TTAAGGGGTATTCTCTGCAAGGAGAAACCTGCTGAAGAGTCCGGCACTGAACCAACACCGG	5817
Db	5761	TTAAGGGGTATTCTCTGCAAGGAGAAACCTGCTGAAGAGTCCGGCACTGAACCAACACCGG	5820
QY	5818	GGCCCGTTCACCTCCCGGAGAGAGCCCAACAGGTGAGGCCACCAAGTACATGACAGAC	5877
Db	5821	GGCCCGTTCACCTCCCGGAGAGAGCCCAACAGGTGAGGCCACCAAGTACATGACAGAC	5880
QY	5878	ATCACCAAGCCGTGGCTCCAGCCCAAGCCGAGCCGCGCAAGGACCCCACTCAACCCGAGAG	5937
Db	5881	ATCACCAAGCCGTGGCTCCAGCCCAAGCCGAGCCGCGCAAGGACCCCACTCAACCCGAGAG	5940
QY	5938	CCAAGACACCCCAACCGCTACCGGAGAGGCGCGACGAGCTCGCGCAGGACAAATCTCTCT	5997
Db	5941	CCAAGACACCCCAACCGCTACCGGAGAGGCGCGACGAGCTCGCGCAGGACAAATCTCTCT	6000
QY	5998	GGCCGCCCTCTGAGCGAGAGAAATCCCGGCGCGGATACTCAGCACGCGGAGAGACCGG	6057
Db	6001	GGCCGCCCTCTCTGAGCGAGAGAAATCCCGGCGCGGATCTCAGCACGCGGAGAGACCGG	6060

Query 6058 TCCCGCGGAGGCTGTTGAAGACAGCAGGAGGCGCGCTGCTGCGGAGCCCTGAGG 6117
Db 6061 TCCCGCGGAGGCTGTTGAAGACAGCAGGAGGCGCGCTGCTGCGGAGCCCTGAGG 6120
Query 6118 ACCCGCGCTGCTCCAGGATGAACAAGCTCTGGGACCAAGTC 6155
Db 6121 ACCCGCGCTGCTCCAGGATGAACAAGCTGAGGACACATTC 6158

RESULT 7
AX503780 6189 bp DNA linear PAT 27-SEP-2002
LOCUS Sequence 10 from Patent WO0226826.
DEFINITION AX503780
ACCESSION AX503780
VERSION AX503780.1 GI:23385965
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 Gerlach, V.L., Macdougall, J.R., Smithson, G., Miller, I., Stone, D.,
Gunter, E., Eilerman, K., Grose, W.M., Alsobrook, J.P., Lepley, D.M.,
Burgess, C.E., Padigaru, M., Kekuda, R., Spytek, K.A., Leach, M.D. and
Shinkens, R.A.
Proteins and nucleic acids encoding same
JOURNAL Patent: WO 0226826-A 10 04-APR-2002;
Curegen Corporation (US)
FEATURES
source Location/Qualifiers
1..6189
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1742 a 1554 c 1690 g 1203 t
ORIGIN

Query Match 99.3%; Score 6121.8; DB 6; Length 6189;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 6145; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

Query 1 ATGTTGAAGTTCAAAATATGAGCGCGGAATCCTTTGATGCTGCTGCTGAACCCATT 60
Db 1 ATGTTGAAGTTCAAAATATGAGCGCGGAATCCTTTGATGCTGCTGCTGAACCCATT 60
Query 61 GCGAGCGGGGCTCCAGGCTGATGCTTCTCCAGGGGAAACCCCTTTATGACTCA 120
Db 61 GCGAGCGGGGCTCCAGGCTGATGCTTCTCCAGGGGAAACCCCTTTATGACTCA 120
Query 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTTGAA 180
Db 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTTGAA 180
Query 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTTGAA 180
Db 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTTGAA 180
Query 181 GAATGCAATCAGCTGCTCTGATGAAGATTAAAGCAGTGAACCTTTGTCGGAAGTAT 240
Db 181 GAATGCAATCAGCTGCTCTGATGAAGATTAAAGCAGTGAACCTTTGTCGGAAGTAT 240
Query 241 TCCGACACCAATGCTGAGTTACAGAGCTCCAGCCTTCCGAAAGAGACTTTGAAAGTCA 300
Db 241 TCCGACACCAATGCTGAGTTACAGAGCTCCAGCCTTCCGAAAGAGACTTTGAAAGTCA 300
Query 301 AGTCTTAGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 301 AGTCTTAGAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Query 361 GACATCTATGCTATGAAGAAGTATGAAGAAGAGCTTTATTTGCCCAGAGAGAGTTTCA 420
Db 361 GACATCTATGCTATGAAGAAGTATGAAGAAGAGCTTTATTTGCCCAGAGAGAGTTTCA 420
Query 421 TTTTATGAG 480
Db 421 TTTTATGAG 480
Query 481 CAGTAGCTTTTCAAGACAAATAACCTTTATCTGCTATGAAATATCAGCCTGAGAGG 540

Db 481 CAGTAGCTTTTCAAGACAAATAACCTTTATCTGCTATGAAATATCAGCCTGAGAGG 540
Query 541 GACTTCTGCTCACTTTTGAATATGATATGAGAGCAGTTAGTAAACCTGATACAGTTT 600
Db 541 GACTTCTGCTCACTTTTGAATATGATATGAGAGCAGTTAGTAAACCTGATACAGTTT 600
Query 601 TACTTAGCTGAGCTGTTTGGCTGTTCAAGGCTTCATCTGATGAGATACGTCATCGA 660
Db 601 TACTTAGCTGAGCTGTTTGGCTGTTCAAGGCTTCATCTGATGAGATACGTCATCGA 660
Query 661 GACATCAAGCTGAGAACATCTCTGTTGACCGGACAGACATCAAGCTGAGATTTT 720
Db 661 GACATCAAGCTGAGAACATCTCTGTTGACCGGACAGACATCAAGCTGAGATTTT 720
Query 721 GATCTGCGCGGAAATGATTTCAACAAAGATGATGCAAACTCCGATGAGGACC 780
Db 721 GATCTGCGCGGAAATGATTTCAACAAAGATGATGCAAACTCCGATGAGGACC 780
Query 781 CCAGATTACATGCTCTGAGAGTCTGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 CCAGATTACATGCTCTGAGAGTCTGATGATGATGATGATGATGATGATGATGATGAT 840
Query 841 GAGCTGAGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 900
Db 841 GAGCTGAGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAG 900
Query 901 TCCCGCTGCGAGAGGAACTCTGCGCAACCTTCATTAATATGATTTCCAGCGG 960
Db 901 TCCCGCTGCGAGAGGAACTCTGCGCAACCTTCATTAATATGATTTCCAGCGG 960
Query 961 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1020
Db 961 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1020
Query 1021 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1080
Db 1021 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1080
Query 1081 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1140
Db 1081 TTTTGAATTTCCAGATGACCCCAAGAGGAGCAGTGAATCTTCTGATGATGATGATGAT 1140
Query 1141 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1141 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1200
Query 1198 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1260
Db 1198 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1260
Query 1261 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1320
Db 1261 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1320
Query 1321 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1380
Db 1321 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1380
Query 1381 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1440
Db 1381 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1440
Query 1441 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1500
Query 1498 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1557
Db 1498 TCTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1557
Query 1561 TTTGAGATGACACCTCCATTTTGAATGATGATGATGATGATGATGATGATGATGAT 1620

Db	1558	TTGGAGCAGCAACGGATGAGAGTGTCGCCAGAGGATGCAAAAGCATGCACTTCTCAT	1617
QY	1621	GATATCAGAGGCGAGAGCCCGGAAGCTCCAAAGATCAAAAGCGAGATCCAGGCTCAA	1680
Db	1618	GATATCAGAGGCGAGAGCCCGGAAGCTCCAAAGATCAAAAGCGAGATCCAGGCTCAA	1677
QY	1681	GTGGAAAGAAATGAGGTTGATGATGATCAATCGTTGGAAAGATCTTGCTCAGCAAGAGA	1740
Db	1678	GTGGAAAGAAATGAGGTTGATGATGATCAATCGTTGGAAAGATCTTGCTCAGCAAGAGA	1737
QY	1741	CGAGATGATCTCTAAGGATCTGAGCTGAGAGATCTGGCTTGGCTCTGAGGATTCAG	1800
Db	1738	CGAGATGATCTCTAAGGATCTGAGCTGAGAGATCTGGCTTGGCTCTGAGGATTCAG	1797
QY	1801	CGGAAGCGAGAGATGTCAGCATMACTGTGAAAGCTAAGGATCAAGAGAAAGCTGAA	1860
Db	1798	CGGAAGCGAGAGATGTCAGCATMACTGTGAAAGCTAAGGATCAAGAGAAAGCTGAA	1857
QY	1861	GTGGAGAGATATGCGAAATCTGAGAAAGTCAATGCTAGCGAGCTCAAAATTCAGAG	1920
Db	1858	GTGGAGAGATATGCGAAATCTGAGAAAGTCAATGCTAGCGAGCTCAAAATTCAGAG	1917
QY	1921	CTCCAAAGAAACTGAGAGAGGCTGTAAAGCCAGCAAGAGGACCTCGAGCTCTGAG	1980
Db	1918	CTCCAAAGAAACTGAGAGAGGCTGTAAAGCCAGCAAGAGGACCTCGAGCTCTGAG	1977
QY	1981	AATATCCGCCACAGGCGAAAGAGCGAGCGAGAGGAGCTGAGAGAGCTGCAAAACCGAG	2040
Db	1978	AATATCCGCCACAGGCGAAAGAGCGAGCGAGAGGAGCTGAGAGAGCTGCAAAACCGAG	2037
QY	2041	GATTCCTTGTAAAGGCGATCAGAAAGAGGCTGTGAAAGCTGAGGAAAGCGCGCATTCCTG	2100
Db	2038	GATTCCTTGTAAAGGCGATCAGAAAGAGGCTGTGAAAGCTGAGGAAAGCGCGCATTCCTG	2097
QY	2101	GAGAACAAAGTTAABAGACTAGAGCCATGAGAGCGTTAAGAAACAGCTBAAGATGAC	2160
Db	2098	GAGAACAAAGTTAABAGAGCTAGAGCCATGAGAGCGTTAAGAAACAGCTBAAGATGAC	2157
QY	2161	ATCCGAGCAAAATCCCAACAGATCCAGCAGATGGCTGATTAATAATTCGAGCTCGAAG	2220
Db	2158	ATCCGAGCAAAATCCCAACAGATCCAGCAGATGGCTGATTAATAATTCGAGCTCGAAG	2217
QY	2221	AAACATCGGGAGGGCCCAAGTCTACGCCAGCACTTAGAAGTGCACCTGAAACAGAAAG	2280
Db	2218	AAACATCGGGAGGGCCCAAGTCTACGCCAGCACTTAGAAGTGCACCTGAAACAGAAAG	2277
QY	2281	CAGCATATGAGGAAAAAGATTAAAGTGTGACAAATCGATTAAGAAAGACTGCGTGAAC	2340
Db	2278	CAGCATATGAGGAAAAAGATTAAAGTGTGACAAATCGATTAAGAAAGACTGCGTGAAC	2337
QY	2341	AAGGAGCACTGCGAACAATGATGAGAGACACCGAGGAGAGGGCCCATGAGAAAGGCGAA	2400
Db	2338	AAGGAGCACTGCGAACAATGATGAGAGACACCGAGGAGAGGGCCCATGAGAAAGGCGAA	2397
QY	2401	ATTCTCAGCCGAAACAGAGGCGATGATCAATGCTATGATATTCACAGATCAGATCCCTGAA	2460
Db	2398	ATTCTCAGCCGAAACAGAGGCGATGATCAATGCTATGATATTCACAGATCAGATCCCTGAA	2457
QY	2461	CAGAGAGATTGAGGAATGCTGTGGAAGCCCAATAACTTGAGCAAAATGACAGCTTTTAAAC	2520
Db	2458	CAGAGAGATTGAGGAATGCTGTGGAAGCCCAATAACTTGAGCAAAATGACAGCTTTTAAAC	2517
QY	2521	CAAAGGAAATGAGAGGCCCAAGAGAGATGATTTCTGAATCTGAGGCAACAGAAATTTTAC	2580
Db	2518	CAAAGGAAATGAGAGGCCCAAGAGAGATGATTTCTGAATCTGAGGCAACAGAAATTTTAC	2577
QY	2581	CTGGAGACACAGGCTGGAGAGTTGAGAGCCCAAGAACCCAAACTGAGAGACAGCTGGAG	2640
Db	2578	CTGGAGACACAGGCTGGAGAGTTGAGAGCCCAAGAACCCAAACTGAGAGACAGCTGGAG	2637
QY	2641	AAGATCAGCCACCAAGACCAACAGTCAAGAATCGGCTGCTGGAATCTGAGACAAAGATTG	2700
Db	2638	AAGATCAGCCACCAAGACCAACAGTCAAGAATCGGCTGCTGGAATCTGAGACAAAGATTG	2697

QY	2701	CGGAGGCGATCTTAGAGCAACGAGACACAGAAATGGAGCTCAAGCGCCAGCTCACAG	276
Db	2698	CGGAGGCGATCTTAGAGCAAGAGAGACAGAAATGGAGCTCAAGCGCCAGCTCACAG	2757
QY	2761	CTACAGCTCTCCCTGCGAGAGCGAGTCAACAGTTGACAGCCCTGCGAGCTGCACGAGCG	2820
Db	2758	CTACAGCTCTCCCTGCGAGAGCGAGATCAAGTTGACAGCCCTGCGAGCTGCACGAGCG	2817
QY	2821	GCCCTGAGAGACCGAGCTTCGCGACGCAAGACABAGCTGGAAGGACCAACGCGAAAGCT	2880
Db	2818	GCCCTGAGAGACCGAGCTTCGCGACGCAAGACABAGCTGGAAGGACCAACGCGAAAGCT	2877
QY	2881	GAAGGAGATCCAGGCACTCACGCACTATGAGATGAATCCAGCGCAATTTGATGCT	2940
Db	2878	GAAGGAGATCCAGGCACTCACGCGCATATGATGAATCCAGCGCAATTTGATGCT	2937
QY	2941	CTTCGTAAACACTGTATCTGTAAATACAGACCTGGAGAGGAGCTTAAACAGCTCACGAG	3000
Db	2938	CTTCGTAAACACTGTATCTGTATACAGACTGGAGAGGAGCTTAAACAGCTCACGAG	2997
QY	3001	GACAACTGTAACTCAACCAACCAAACTTCTACTGTTCAAACAACCTGATGAGGCTTCT	3060
Db	2998	GACAACTGTAACTCAACCAACCAAACTTCTACTGTTCAAACAACCTGATGAGGCTTCT	3057
QY	3061	GCGCGCAACGACGAGATTGTACACTGGAGATGAAGGACCAATTCGCGCGGAGATC	3120
Db	3058	GCGCGCAACGACGAGATTGTCAACTGCGAGATGAAGGACCAATTCGCGCGGAGATC	3117
QY	3121	ACGGAAAGAGATGACAGCTTACCAACCCAGAAAGCAACGATGAGGCTCTGAGACCAACG	3180
Db	3118	ACGGAAAGAGATGACAGCTTACCAACCCAGAAAGCAACGATGAGGCTCTGAGACCAACG	3177
QY	3181	TGCACACATGCTGGAGGAAACAGGTACATGATTTGAGAGCCCTAAACGATGACCTCTGAA	3240
Db	3178	TGCACACATGCTGGAGGAAACAGGTATGATTTGAGAGCCCTAAACGATGACCTCTGAA	3237
QY	3241	AAAGAGCGGACGTGGAGGCTTGGAGGACGCTCTGGGTGAATGAGAAATCCCACTTTGAG	3300
Db	3238	AAAGAGCGGACGTGGAGGCTTGGAGGACGCTCTCTGGGTGAATGAGAAATCCCACTTTGAG	3297
QY	3301	TGTGGGTTCAGAGCTGCGACAGATGCTGGACACCGAGAAACAGAGCGGCGAGAGCC	3360
Db	3298	TGTGGGTTCAGAGCTGCGAGAGATGCTGGACACCGAGAAACAGAGCGGCGAGAGCC	3357
QY	3361	GATCAGCGATCACCGAGCTCTGCGAGGTGGAGCTGCGAGTGAAGGACCAACAGGCT	3420
Db	3358	GATCAGCGATCACCGAGCTCTGCGAGGTGGAGCTGCGAGTGAAGGACCAACAGGCT	3417
QY	3421	GAGATTCCTGCTCTGCGACAGGCTCTCAAACAGCAGAACTGGAAGGCGGAGGCTCTCT	3480
Db	3418	GAGATTCCTGCTCTGCGACAGGCTCTCAAACAGCAGAACTGGAAGGCGGAGGCTCTCT	3477
QY	3481	GACAGCTCAATGACTCTGAGAGAAAGCATGCTATGCTGAATGAATGCCCGAAGCTTGA	3540
Db	3478	GACAGCTCAATGACTCTGAGAGAAAGCATGCTATGCTGAATGAATGCCCGAAGCTTGA	3537
QY	3541	CAGCAGAGCTGAGAGCTGGAACGAGAGCTCAACAGAGGCTTCTGGAAGACAGCCAAA	3600
Db	3538	CAGCAGAGCTGAGAGCTGGAACGAGAGCTCAACAGAGGCTTCTGGAAGACAGCCAAA	3597
QY	3601	TTACAGCAGCAGATGAGCTTGCAAGAAAATCAATTTCCGTCTGACTCAAGACTGCA	3660
Db	3598	TTACAGCAGCAGATGAGCTTGCAAGAAAATCAATTTCCGTCTGACTCAAGACTGCA	3657
QY	3661	GAAGCTCAGATCGGCTGATCTACTGAAAGCAAGAAAGATGACTTGAGATACAGCTG	3720
Db	3658	GAAGCTCAGATCGGCTGATCTACTGAAAGCAAGAAAGATGACTTGAGATACAGCTG	3717
QY	3721	GAAGCAATTCAGGTTCTTATTTCTCATGAAAGGTGAATGAAAGCACTATTTCTGAA	3780
Db	3718	GAAGCAATTCAGGTTCTTATTTCTCATGAAAGGTGAATGAAAGCACTATTTCTGAA	3777

OY	3781	AAACCAACATCATGATTTTCTGTGACACCAAAATGACCAACCTGCTAAAGAAAAAG	3840
Db	3778	CAACCAAACTCATTTGATTTTCTGTACAGCCAAATGACCAACTGCTAAAGAAAAAG	3837
OY	3841	GTTCCCTGCACTAGCAATGAAGCTGAGCTGAGCCCTGTAGAGAGAGAAAGTCTGTGTGCA	3900
Db	3838	GTGCCTCTGCAGTACATGATGAGTGAACTGGGCCCTGGAGAGAGAGAAAGCTCGGTGTGCA	3897
OY	3901	GAGCTTAGAGAAAGCCCTTCAGAAAGACCCCGCATCGAGCTCCGGGTCCGGGAGAGAAAGCT	3960
Db	3898	GAGCTTAGAGAGAGCCCTTCAGAAAGACCCCGCATCGAGCTCCGGGTCCGGGAGAGAAAGCT	3957
OY	3961	GCCCAACGCAAGCAACGAGCAACCAACCAACCCATCTCAAGCCAGCAACCCGAGGAGACAG	4020
Db	3958	GCCCAACGCAAGCAACGAGCAACCAACCAACCCATCTCAAGCCAGCAACCCGAGGAGACAG	4017
OY	4021	ATCCGCAATGTCCGCAATCTGTGCGGTCCGCAAGACACAGCCCAAGTCCATGAGCTGTGTG	4080
Db	4018	ATCCGCAATGTCTGGCATGTGTGCGGTCCGCAAGACACAGCCCAAGTCCATGAGCTGTGTG	4077
OY	4081	GCCCCGCAATCAAGCCGAGAAAGAGTCTCAACTCCAGAGAAATTAGTGGAGCTT	4140
Db	4078	GCCCCGCAATCAAGCCGAGAAAGAGTCTCTCACTCCAGAGAAATTAGTGGAGCTT	4137
OY	4141	AAGAAACGCATGACCAACATATTCTCAACGATTCAACGTAGACGTGAACATCGAGCC	4200
Db	4138	AAGAAACGCATGACCAACATATTCTCAACGATTCAACGTAGACGTGAACATCGAGCC	4197
OY	4201	ACAAAGT	4260
Db	4198	ACAAAGT	4257
OY	4261	GAATGTCAAGT	4320
Db	4258	GAATGTCAAGT	4317
OY	4321	CCTGCTGAATATGCAACACATTTCAACGAGGCTTTTGTGCTGTGCAAAATGAATCTCCCA	4380
Db	4318	CCTGCTGAATATGCAACACATTTCAACGAGGCTTTTGTGCTGTGCAAAATGAATCTCCCA	4377
OY	4381	GGTCTCCGACCAAGAGACCAGACGACGTGTGACCTGGAAGGTTGATGAAGTGTGCC	4440
Db	4378	GGTCTCCGACCAAGAGACCAGACGACGTGTGACCTGGAAGGTTGATGAAGTGTGCC	4437
OY	4441	AGGAATTAACAACGAGACAGCAAGGCTGTGAGACAGAAATACATTTGCTTGAAGGATCA	4500
Db	4438	AGGAATTAACAACGAGACAGCAAGGCTGTGAGACAGAAATACATTTGCTTGAAGGATCA	4497
OY	4501	AAAGTCCCATTTTATGACATGAAGCCGAGAGCTGTGACAGAGGCCGTGTGAAGAAATT	4560
Db	4498	AAAGTCCCATTTTATGACATGAAGCCGAGAGCTGTGACAGAGGCCGTGTGAAGAAATT	4557
OY	4561	GAGCTGTGCTTCCGACGAGGAGTGTATCTATGATGTGTGCTGTGTGTGTGTGTGTGTGT	4620
Db	4558	GAGCTGTGCTTCCGACGAGGAGTGTATCTATGATGTGTGCTGTGTGTGTGTGTGTGTGT	4617
OY	4621	GCAAAATNAGGCAAGACGATGTCCCAATACATGAAATGAGATCTCACCCGCAACAC	4680
Db	4618	GCAAAATNAGGCAAGACGATGTCCCAATACATGAAATGAGATCTCACCCGCAACAC	4677
OY	4681	ACCTGCTGTGACCGGAGAAACCTCTACTTGTGTAGTCCCACTTCCCTGACAAACAGGAC	4740
Db	4678	ACCTGCTGTGACCGGAGAAACCTCTACTTGTGTAGTCCCACTTCCCTGACAAACAGGAC	4737
OY	4741	TGGGTCAACGCGCTTAGAATCAAGTTGTGCAAGTGTGAGAGATTTCTTAGGAAAAAGCAGAA	4800
Db	4738	TGGGTCAACGCGCTTAGAATCAAGTTGTGCAAGTGTGAGAGATTTCTTAGGAAAAAGCAGAA	4797
OY	4801	GCTGATGTCTAACTGCTGTGAAACCTCCCTGTGTGAACTGGAAGGTGATGACGCTGTGAC	4860
Db	4798	GCTGATGTCTAACTGCTGTGAAACCTCCCTGTGTGAACTGGAAGGTGATGACGCTGTGAC	4857
OY	4861	ATGAATCTGCACGCTGCTTCACTGATCAACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4920

Db	4858	ATGAAC	TGCAAGCTG	CCCTTC	ACGACCAAGT	ATGTTGGTGGCACC	GAGAAAGGCTC	4917	
QY	4921	TAGC	CCCTGAATG	CTTGA	AAAACTCCCTA	ACCATGTCC	CGAATTTGAGCAGTCTTC	4980	
Db	4918	TACG	CCCTGAATG	CTTGA	AAAACTCCCTA	ACCATGTCC	CGAATTTGAGCAGTCTTC	4977	
QY	4991	CAAA	TTTATATAT	CAAGAG	CCCTG	BAAGTACTCAT	GTATAGAGAGAAAGACCGGCA	5048	
Db	4978	CAAA	TTTATATAT	CAAGAG	CCCTG	BAAGTACTCAT	GTATAGAGAGAAAGACCGGCA	5037	
QY	5041	CTGT	GCTTTGAGCAGTGA	AAAAAGT	AAACAGTCC	TGGCCAGTCC	CACTGCTGCTCC	5100	
Db	5038	CTGT	GCTTTGAGCAGTGA	AAAAAGT	AAACAGTCC	TGGCCAGTCC	CACTGCTGCTCC	5099	
QY	5101	CAG	CCCGACATCTCA	CCCA	CATT	TTTGAAGCTGT	CAAGGGCTGCCA	CTTGTTTGGGGCA	5166
Db	5098	CAG	CCCGACATCTCA	CCCA	CATT	TTTGAAGCTGT	CAAGGGCTGCCA	CTTGTTTGGGGCA	5157
QY	5181	GGCA	AAGATTGGA	AOGGGCT	GTGATCTGT	GTGACACGANT	GCCACGAAAGTGTCTATCTC	5220	
Db	5158	GGCA	AAGATTGGA	AOGGGCT	GTGATCTGT	GTGACACGANT	GCCACGAAAGTGTCTATCTC	5217	
QY	5221	CGCT	CAACGAAAACT	CAGCAAA	TACTGCAT	CCGGA	AAAGATAGAGACCTCAGAGCC	5280	
Db	5218	CGCT	CAACGAAAACT	CAGCAAA	TACTGCAT	CCGGA	AAAGATAGAGACCTCAGAGCC	5277	
QY	5281	TGCA	GCTGTATCCACCTT	CACCA	ATTACAGTAT	CCTCAT	TGTGGAACCATTAATTTCTAGCA	5340	
Db	5278	TGCA	GCTGTATCCACCTT	CACCA	ATTACAGTAT	CCTCAT	TGTGGAACCATTAATTTCTAGCA	5337	
QY	5341	ATCG	ACATGAAAGCAGTAC	AGCCTCG	AGAGAA	TTCTGTG	ATAGAAATGACATTTCTTGCA	5400	
Db	5338	ATCG	ACATGAAAGCAGTAC	AGCCTCG	AGAGAA	TTCTGTG	ATAGAAATGACATTTCTTGCA	5397	
QY	5401	CTGT	CTGTGTTTGC	CGGCTCTT	CCAACA	AGCTTCC	CTGTCTCAATCTGTGACGTGAACAGC	5460	
Db	5398	CTGT	CTGTGTTTGC	CGGCTCTT	CCAACA	AGCTTCC	CTGTCTCAATCTGTGACGTGAACAGC	5457	
QY	5461	GCAG	GGCAGCAGAGAGAGT	ACTTGT	CTGTGTTTCA	CAGAA	TTTGGAGTGTGTGGAATTC	5520	
Db	5458	GCAG	GGCAGCAGAGAGAGT	ACTTGT	CTGTGTTTCA	CAGAA	TTTGGAGTGTGTGGAATTC	5517	
QY	5521	TACG	GAAGCCTAGCCG	CAAGAC	GATCTCA	AGTGA	GTGTGCTTACCTTTGGCTTTGCC	5580	
Db	5518	TACG	GAAGCCTAGCCG	CAAGAC	GATCTCA	AGTGA	GTGTGCTTACCTTTGGCTTTGCC	5577	
QY	5581	TACG	AGAAACCTTATCT	GTGTTGT	ATACCA	CTTCA	CACTCGAAGTATTTGAAATCCAG	5640	
Db	5578	TACG	AGAAACCTTATCT	GTGTTGT	ATACCA	CTTCA	CACTCGAAGTATTTGAAATCCAG	5637	
QY	5641	GCA	GCTCTTCAGAGAGG	AGCCCT	TGCCAG	CGTACT	GTGACATCCCGAAACCGCGCTAC	5700	
Db	5638	GCA	GCTCTTCAGAGAGG	AGCCCT	TGCCAG	CGTACT	GTGACATCCCGAAACCGCGCTAC	5697	
QY	5701	CTGG	CCCTGCGCATTTT	CTCTCA	GAGCGCATTT	ACTTGG	CGTCTCATATCAGAGTAATTA	5760	
Db	5698	CTGG	CCCTGCGCATTTT	CTCTCA	GAGCGCATTT	ACTTGG	CGTCTCATATCAGAGTAATTA	5757	
QY	5761	AGGT	CAATTTTCTGCAAGG	AAAACT	CTGTGTA	AGGATCC	GGCACTGAACACACCGGGGC	5820	
Db	5758	AGGT	CAATTTTCTGCAAGG	AAAACT	CTGTGTA	AGGATCC	GGCACTGAACACACCGGGGC	5817	
QY	5821	CGGT	CACTTCGCGAGAG	AGCCCA	CAAGCGAG	GCCCAT	CCCATGTAACAGACACATC	5880	
Db	5818	CGGT	CACTTCGCGAGAG	AGCCCA	CAAGCGAG	GCCCAT	CCCATGTAACAGACACATC	5877	
QY	5881	ACCA	AGCGCGTGGCTC	ACAGCC	CAAGGCGCG	CCGGA	AGGCCACACCCCGGAGAGCA	5940	
Db	5878	ACCA	AGCGCGTGGCTC	ACAGCC	CAAGGCGCG	CCGGA	AGGCCACACCCCGGAGAGCA	5937	
QY	5941	AGCA	CACCCACCGCTA	CCGCGAG	GGCGGAC	CGAGCTG	CGCAGGGA	CAAGTCTCTGGC	6000

Db 5938 AGACACCCACCGCTACCGGAGGGGCGACCGAGTGGCGACAGGACAAAGTCTCTGAC 5997
QY 6001 CGCCCCCTGAGCGGAGAGAGTCCCCCGGCGGATATCTGACACCGCGAGAGCGGTCC 6060
Db 5998 CGCCCCCTGAGCGGAGAGAGTCCCCCGGCGGATATCTGACACCGCGAGAGCGGTCC 6057
QY 6061 CCCGCGAGGCTGTTTGAAGACAGACGAGGCGCGCTGCTGCGGAGCGGTGAGAAC 6120
Db 6058 CCCGCGAGGCTGTTTGAAGACAGACGAGGCGCGCTGCTGCGGAGCGGTGAGAAC 6117
QY 6121 CCGCTGTCCAGGTGAACAAAGTCTGGGACCACTTCAATATTA 6165
Db 6118 CCGCTGTCCAGGTGAACAAAGTCTGGGACCACTTCAATATTA 6162

RESULT 8
AX503778
LOCUS AX503778 6201 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 8 from Patent WO226826.
ACCESSION AX503778
VERSION AX503778.1 GI:23385964
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1
AUTHORS Gelach, V.L., Macdougall, J.R., Smithson, G., Miller, I., Stone, D.,
Gunter, E., Ellerman, K., Grosse, W.M., Alsbrock, J.P., Lempy, D.M.,
Burgess, C.E., Padigar, M., Kekuda, R., Spyrek, K.A., Leach, M.D. and
Shimkets, R.A.
TITLE Proteins and nucleic acids encoding same
JOURNAL Patent: WO 0226826-A 8 04-APR-2002;
Curagen Corporation (US)
FEATURES
source 1..6201
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 1736 a 1552 c 1707 g 1206 t
ORIGIN

Query Match 99.0%; Score 6105.4; DB 6; Length 6201;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 6131; Conservative 0; Mismatches 21; Indels 3; Gaps 1;

QY 1 ATGTTGAAGTTCAATATGAGCGCGGAATCTTTGATGCTGCTGTAACCAT 60
Db 1 ATGTTGAAGTTCAATATGAGCGCGGAATCTTTGATGCTGCTGTAACCAT 60
QY 61 GCCAGCGGGCTCCAGGCTGATCTCTTCCAGAGGAACACCCCTTTATGACTCA 120
Db 61 GCCAGCGGGCTCCAGGCTGATCTCTTCCAGAGGAACACCCCTTTATGACTCA 120
QY 121 CAGCAGATGTCCTCTTCCAGAGGAGATATGATGCCCTCTTTCTTTGAA 180
Db 121 CAGCAGATGTCCTCTTCCAGAGGAGATATGATGCCCTCTTTCTTTGAA 180
QY 181 GAATGCAAGTACGCTCTCTGATGAAGATTAAAGCACTTTGTCGGAAT 240
Db 181 GAATGCAAGTACGCTCTCTGATGAAGATTAAAGCACTTTGTCGGAAT 240
QY 241 TCCGACACCATAGCTAGTACAGAGCTCCAGCTTCGCGAAAGGACTTCAGAGTCAGA 300
Db 241 TCCGACACCATAGCTAGTACAGAGCTCCAGCTTCGCGAAAGGACTTCAGAGTCAGA 300
QY 301 AGTCTTGATGTTGTGTGTCATTTGCTGAAGTGCAGTGTGTAAGAGAAACCAACCGGG 360
Db 301 AGTCTTGATGTTGTGTGTCATTTGCTGAAGTGCAGTGTGTAAGAGAAACCAACCGGG 360
QY 361 GACATCTATGCTATGAAAGATGAAGAGAGGCTTTATGGCCCGAGAGAGAGTTTCA 420
Db 361 GACATCTATGCTATGAAAGATGAAGAGAGGCTTTATGGCCCGAGAGAGAGTTTCA 420

QY 421 TTTTGTGAGAGAGCGGAACATATATCTGTGAAGCACAGCCGTGATCCCCCATTA 480
Db 421 TTTTGTGAGAGAGCGGAACATATATCTGTGAAGCACAGCCGTGATCCCCCATTA 480
QY 481 CAGTATGCTTTACAGACAAAATACCTTTATCTGGTCATGGAATATCACTCGAGGG 540
Db 481 CAGTATGCTTTACAGACAAAATACCTTTATCTGGTCATGGAATATCACTCGAGGG 540
QY 541 GACTTGCTGTCACTTTGAATATGATGAGACGATGATGAAACCTGATACAGTTT 600
Db 541 GACTTGCTGTCACTTTGAATATGATGAGACGATGATGAAACCTGATACAGTTT 600
QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCACAGCTTCACTGATGAGATACGTGATCG 660
Db 601 TACCTAGCTGAGCTGATTTTGGCTGTTCACAGCTTCACTGATGAGATACGTGATCG 660
QY 661 GACATCAAGCTGAGAACATCTGTTGACCGCACAGACACATCAAGCTGATGATTTT 720
Db 661 GACATCAAGCTGAGAACATCTGTTGACCGCACAGACACATCAAGCTGATGATTTT 720
QY 721 GATCTGCGCGGAAATATGATCAACAGATGTAATGCAATCTCCGATGGAGC 780
Db 721 GATCTGCGCGGAAATATGATCAACAGATGTAATGCAATCTCCGATGGAGC 780
QY 781 GATCTGCGCGGAAATATGATCAACAGATGTAATGCAATCTCCGATGGAGC 777
Db 781 GATCTGCGCGGAAATATGATCAACAGATGTAATGCAATCTCCGATGGAGC 777
QY 778 CCAAGTTACATGCTCTGAGAGCTGCTGATGTAACCGGAGATGAAAGGACCTAC 840
Db 778 CCAAGTTACATGCTCTGAGAGCTGCTGATGTAACCGGAGATGAAAGGACCTAC 840
QY 841 GGCCTGAGCTGATGCTGTGATGATGAGGCTGATGATGATGATGATGATGATGATG 900
Db 841 GGCCTGAGCTGATGCTGTGATGATGAGGCTGATGATGATGATGATGATGATGATG 900
QY 901 TCCCTGCTGAGAGGAGAACTCTGCGGAGAACTCTGAGAACTCTGAGAACTCTGAG 960
Db 901 TCCCTGCTGAGAGGAGAACTCTGCGGAGAACTCTGAGAACTCTGAGAACTCTGAG 960
QY 961 TTTTGAATTTCCAGATGACCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020
Db 961 TTTTGAATTTCCAGATGACCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020
QY 1021 TTTTGAATTTCCAGATGACCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1080
Db 1021 TTTTGAATTTCCAGATGACCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1080
QY 1078 TCTAAATTTGCTGGAACAACTCTGTAAGTCTCTCCCTGCTGCTGCTGCTGCTG 1137
Db 1078 TCTAAATTTGCTGGAACAACTCTGTAAGTCTCTCCCTGCTGCTGCTGCTGCTG 1137
QY 1141 TCTGACATGACACCTTCAATTTGATGAACCAAGAGAAATTCGAGTTTCACTCT 1200
Db 1141 TCTGACATGACACCTTCAATTTGATGAACCAAGAGAAATTCGAGTTTCACTCT 1200
QY 1201 CCGTGGCAGCTGAGACCCCTGAGGCTCTCGGGGTGAAGAACTGCGGTTTGGGGTTTTCG 1260
Db 1201 CCGTGGCAGCTGAGACCCCTGAGGCTCTCGGGGTGAAGAACTGCGGTTTGGGGTTTTCG 1260
QY 1258 TACAGCAAGGCACTGAGGATCTTGTGATGATGATGATGATGATGATGATGATG 1317
Db 1258 TACAGCAAGGCACTGAGGATCTTGTGATGATGATGATGATGATGATGATGATG 1317
QY 1321 CCGTCAAGATAGCTGATGAGAAAGAACTTCTCAAAAGCAAGGCTGATCAAGAC 1380
Db 1321 CCGTCAAGATAGCTGATGAGAAAGAACTTCTCAAAAGCAAGGCTGATCAAGAC 1380
QY 1378 TCTCAGAGCAAGTGTCAAGATGAGCAGAGATGACCGGTTACATCGAGAGTGTCA 1437
Db 1378 TCTCAGAGCAAGTGTCAAGATGAGCAGAGATGACCGGTTACATCGAGAGTGTCA 1437
QY 1441 GAGGTGAGAGGTGTGCTTATGATGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 1500
Db 1441 GAGGTGAGAGGTGTGCTTATGATGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 1500
QY 1497 GAGGTGAGAGGTGTGCTTATGATGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 1557
Db 1497 GAGGTGAGAGGTGTGCTTATGATGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 1557

QY	1501	TCCTCCCTGGAGCAGAGACCTTGCTCACTCAATCAACAGAAATCAGTAGCTTTAAAGGAAGT	1560
Db	1498	TCCCTCCCTGGAGCAGAGACCTTGCTCACTCAATCAACAGAAATCAGTAGCTTTAAAGGAAGT	1557
QY	1561	TTGGAGCAGAGCCGATGAGGTGTCCAGAGAGATGACAAGACCTGCAGCTTCCAT	1620
Db	1558	TTGGAGCAGAGCCGATGAGGTGTCCAGAGAGATGACAAGACCTGCAGCTTCCAT	1617
QY	1621	GATATCAAGAGCAGAGCCGAGAGCTCCAGAAATCAAGAGCAGAGATACAGGCTCA	1680
Db	1618	GATATCAAGAGCAGAGCCGAGAGCTCCAGAAATCAAGAGCAGAGATACAGGCTCA	1677
QY	1681	GTGGAGAAATCAGGTTGATGATGATCAATCAATTGAGAGATCTTGCTCAGACAAGAGA	1740
Db	1678	GTGGAGAAATCAGGTTGATGATGATCAATCAATTGAGAGAGATCTTGCTCAGACAAGAGA	1737
QY	1741	CGAGATGATCTCTACGAATCTGAGCTGAGAGATCTCGGCTTGCTGCTGAAGATTCAG	1800
Db	1738	CGAGATGATCTCTACGAATCTGAGCTGAGAGATCTCGGCTTGCTGCTGAAGATTCAG	1797
QY	1801	CGAAGAGCAGACAGATGTCAACATTAATCTGTTAAGGCTAAGATCAAGGAGAGGCTGAA	1860
Db	1798	CGAAGAGCAGACAGATGTCAACATTAATCTGTTAAGGCTAAGATCAAGGAGAGGCTGAA	1857
QY	1861	GTGGAGAAATTCGAAACTGAGAGATCAATGCTGAGCAGCAGCTCAAAATTCAGAG	1920
Db	1858	GTGGAGAAATTCGAAACTGAGAGATCAATGCTGAGCAGCAGCTCAAAATTCAGAG	1917
QY	1921	CTCCAAAGAGAACTGAGAGAGGCTGTAAAGCAGACAGAGGCGCACAGACCTGTGAG	1980
Db	1918	CTCCAAAGAGAACTGAGAGAGGCTGTAAAGCAGACAGAGGCGCACAGACCTGTGAG	1977
QY	1961	AATATCGCCACAGGCCAAGAGCGAGCCGAGAGGAGCTGAGAACTTCAGAACCGAG	2040
Db	1978	AATATCGCCACAGGCCAAGAGCGAGCCGAGAGGAGCTGAGAACTTCAGAACCGAG	2037
QY	2041	GATCTCTGTGAGGATCAAGAAAGAGCTGTGGAAGCTGAGGAAGCGCGCATCTCTG	2100
Db	2038	GATCTCTGTGAGGATCAAGAAAGAGCTGTGGAAGCTGAGGAAGCGCGCATCTCTG	2097
QY	2101	GAGAAACAAGTAAAGAGACTGAGAGCCATGAGAGCTAAGAAAACAGACTAAGAGATGAC	2160
Db	2098	GAGAAACAAGTAAAGAGACTGAGAGCCATGAGAGCGTAAAGAAAACAGACTAAGAGATGAC	2157
QY	2161	ATCCAGATAAATCCCAACAGATCCAGAGATGCTGTAAATTTGAGACTCGAAG	2220
Db	2158	ATCCAGATAAATCCCAACAGATCCAGAGATGCTGTAAATTTGAGACTCGAAG	2217
QY	2221	AAACATCGGAGGCCCAAGCTCTCAGCCAGACCTAGAAGTGCACCTGAACACAGAAAG	2280
Db	2218	AAACATCGGAGGCCCAAGCTCTCAGCCAGACCTAGAAGTGCACCTGAACACAGAAAG	2277
QY	2281	CAGCATATGAGGAAGATTTAAAGTGTGGAACAATCAGATTAAGAAAGACCTGCTGAC	2340
Db	2278	CAGCATATGAGGAAGATTTAAAGTGTGGAACAATCAGATTAAGAAAGACCTGCTGAC	2337
QY	2341	AAGGAGCAGCTGGAGAAACATGATGAGAGACAGCAGAGAGAGGCCCATGAGAGGGCA	2400
Db	2338	AAGGAGCAGCTGGAGAAACATGATGAGAGACAGCAGAGAGAGGCCCATGAGAGGGCA	2397
QY	2401	ATTCTCAGCGAACAGAAAGCGCATGATCAATGTGATTCAGAGATCAGATCCCTGGA	2460
Db	2398	ATTCTCAGCGAACAGAAAGCGCATGATCAATGTGATTCAGAGATCAGATCCCTGGA	2457
QY	2461	CAGAGGATGAGGAATCTGTGAGAGCCAAATTAACCTGAGAGAAATPAGAGCTTTTAC	2520
Db	2458	CAGAGGATGAGGAATCTGTGAGAGCCAAATTAACCTGAGAGAAATPAGAGCTTTTAC	2517
QY	2521	CAAGAGAAATGAGAGGCCCAAGAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC	2580
Db	2518	CAAGAGAAATGAGAGGCCCAAGAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC	2577
QY	2581	CTGAGAGACAGGCTGGGAAGTTGAGAGCCAGAAACCGAAATCTGAGAGGACAGCTGAG	2640

Db	2578	CTGAGACACAGAGCTGGGAAGTTGAGAGCCGACGACCGAATACTGAGAGAGCAGCTGGAG	2637
OY	2641	AAGATCAGCCACCAAGACCAACAGTGCACAGAAATCGCTGCTGAACTGAGACAGATTG	2700
Db	2638	AAGATCAGCCACCAAGAGCCACAGTGCACAGAAATCGCTGCTGAACTGAGACAGATTG	2697
OY	2701	CGGAGAGGTCAAGTCTGAGAGACCGAGGAGACCAAAATCGAGCTCAAGGCGCAGCTCACAGAG	2760
Db	2698	CGGAGAGGTGAGTCTGAGAGACCGAGGAGACCAAAATCGAGCTCAAGGCGCAGCTCACAGAG	2757
OY	2761	CTACAGCTCTCCCTGACGAGCGCGAGTCA CAGTTGCACAGCCCTGCGAGGCTGACCGGCG	2820
Db	2758	CTACAGCTCTCCCTGACGAGCGCGAGTCA CAGTTGCACAGCCCTGCGAGGCTGACCGGCG	2817
OY	2821	GCCTTGAGAGCCAGCTTCCGCGAGCGAAGACAGAGCTGAGAGATCCACGCAAGACT	2880
Db	2818	GCCTTGAGAGCCAGCTTCCGCGAGCGAAGACAGAGCTGAGAGATCCACGCAAGACT	2877
OY	2881	GAAAGAGAGATCCAGGCACTCACGGGCATAGAGATAAATCCAGCCCAATTGATGCT	2940
Db	2878	GAAAGAGAGATCCAGGCACTCACGGGCATAGAGATAAATCCAGCCCAATTGATGCT	2937
OY	2941	CTTCGTAAACAGCTGTACTGTAAATCA CAGACTTGAGAGGAGCTTAAACAGCTGACCGAG	3000
Db	2938	CTTCGTAAACAGCTGTACTGTGTATCA CAGACTTGAGAGGAGCTTAAACAGCTGACCGAG	2997
OY	3001	GACAAAGCTGAATCAACCAACCAAAATTTCTACTGTCCAACCAACTCGATGAGGCTCT	3060
Db	2998	GACAAAGCTGAATCAACCAACCAAAATTTCTACTGTCCAACCAACTCGATGAGGCTCT	3057
OY	3061	GGCGCCAACGACAGATTTGTAAACTGCGAAGTGAATGAGCA TCTCCGCGCGGAGATC	3120
Db	3058	GGCGCCAACGACAGATTTGTAAACTGCGAAGTGAATGAGCA TCTCCGCGCGGAGATC	3117
OY	3121	ACGGAACGAGAGATGCGACGCTTACACGACCAAGCAAA CAGTGAAGGCTCTGAGACCAACG	3180
Db	3118	ACGGAACGAGAGATGCGACGCTTACACGACCAAGCAAA CAGTGAAGGCTCTGAGACCAACG	3177
OY	3181	TGCACCAATGCTGAGAGAA CAGGTCA TGGATTTGAGGCGCTTAAC CAGTGAAGCTCTTGAA	3240
Db	3178	TGCACCAATGCTGAGAGAA CAGGTCA TGGATTTGAGGCGCTTAAC CAGTGAAGCTCTTGAA	3237
OY	3241	AAAGAGCGGACGTGGAGGCTTGAGAGGCGTCTGCTGATGAGAAATCCAGTTTGAG	3300
Db	3238	AAAGAGCGGACGTGGAGGCGCTTGAGAGGCGTCTGCTGATGAGAAATCCAGTTTGAG	3297
OY	3301	TGTCGGGTTGAGAGCTGCGAGAGGATCTGGA CACCGAGAAACAGACAGGCGGAGAGCC	3360
Db	3298	TGTCGGGTTGAGAGCTGCGAGAGGATCTGGA CACCGAGAAACAGACAGGCGGAGAGCC	3357
OY	3361	GATCAGCGGATCACCGAGTCTTGCCAGAGTGTGAGCTGGCAGTGAAGAGACCAAGCT	3420
Db	3358	GATCAGCGGATCACCGAGTCTTGCCAGAGTGTGAGCTGGCAGTGAAGAGACCAAGCT	3417
OY	3421	GAGATTCCTGCTGCGACGAGCTCTCAAGAGAGAGAGCTGAAGGCGCGAGGCTCTCT	3480
Db	3418	GAGATTCCTGCTGCGACGAGCTCTCAAGAGAGAGAGCTGAAGGCGCGAGGCTCTCT	3477
OY	3481	GACAAAGCTCAATGACCTGAGAGAAAGCATGCTATGCTTAAATGATGCGCCGAGCTT	3540
Db	3478	GACAAAGCTCAATGACCTGAGAGAAAGCATGCTATGCTTAAATGATGCGCCGAGCTT	3537
OY	3541	CAGAGAAAGCTGAGAGATCTGAACGAGAGCTCAAC CAGAGGCTTCTGGAAGAGCAAGCTAAA	3600
Db	3538	CAGAGAAAGCTGAGAGATCTGAACGAGAGCTCAAC CAGAGGCTTCTGGAAGAGCAAGCTAAA	3597
OY	3601	TTACAGACGACGATGAGCTGCGAGAAATAATCAATTTTCCGTCTGACTCAAGAGACTGCA	3660
Db	3598	TTACAGACGACGATGAGCTGCGAGAAATAATCAATTTTCCGTCTGACTCAAGAGACTGCA	3657
OY	3661	GAAAGCTTAAGATGGGCTGATCTTACTGAAAGACGAAAGAAATGACTTGGAGTCAAGCTG	3720

QY 5881 ACCAAGCGGTGCTCCAGCCAGCCGCGCCGCGGAGCCGAGCCAGCCGAGAGCCCA 5940
 DB 5878 ACCAAGCGGTGCTCCAGCCAGCCGCGCCGCGGAGCCGAGCCAGCCGAGAGCCCA 5937
 QY 5941 AGACACACCCACCGCTACCGCGAGGAGGAGCCAGGCTCGGAGGAGCAAGTCTCTGAC 6000
 DB 5938 AGACACACCCACCGCTACCGCGAGGAGGAGCCAGGCTCGGAGGAGCAAGTCTCTGAC 5997
 QY 6001 CGCCCGCTGAGCGAGAGAGTCCCGCGGCGGATACAGCAGCGGAGAGAGGAGTCC 6060
 DB 5998 CGCCCGCTGAGCGAGAGAGTCCCGCGGCGGATACAGCAGCGGAGAGAGGAGTCC 6057
 QY 6061 CCGGAGGCTGTTTGAAGACAGCAGGAGGCGCGGCTCGGAGGAGCGGTGAGAGCC 6120
 DB 6058 CCGGAGGCTGTTTGAAGACAGCAGGAGGCGCGGCTCGGAGGAGCGGTGAGAGCC 6117
 QY 6121 CCGGCTGCTCCAGGTGACAGAGGTCTGGGACGAGTCC 6155
 DB 6118 CCGGCTGCTCCAGGTGACAGAGGTGAGGAGGAGTCC 6152

RESULT 9
 LOCUS AY257469 8576 bp mRNA linear PRI 23-APR-2003
 DEFINITION Homo sapiens rho/rac-interacting citron kinase (CIT) mRNA, complete cds.
 ACCESSION AY257469
 VERSION AY257469.1 GI:30088969
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Huang, C.Q., Wu, S.L., Shan, Y.X., Liu, S. and Xiao, P.J.
 TITLE Direct Substitution
 JOURNAL Submitted (18-MAR-2003) Department of Biochemistry and Molecular Biology, Basic Medicine, Suzhou University, Renming Road 48, Suzhou, Jiangsu 215007, China

FEATURES

source 1..8576
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12q24.23"
 1..8576
 /gene="CIT"
 /note="synonym: CRK"
 54..6137
 /gene="CIT"
 /note="serine/threonine protein kinases"
 /codon_start=1
 /product="rho/rac-interacting citron kinase"
 /db_xref="GI:30088970"
 /translation="MLKKRYGARNPLDGAAPFASBRNLNLFQGRPPMTQOMS
 PLSRSLDLATVLFEBGSPALMKIKHNSFNKYSPTIAELQJLSAKDFEKL
 VGGFAEVQVVEKATSDIYAMKWKVALLAOQVQSFEEERNILSRSPMPOL
 QYAFQDKNLHYLMVEYQPGDLSLNRVEDQDNLIOFYLAELILVSHVHMGY
 HDIKENILVDRTHIKLDFGSAKMSNPNVAKLPIGTPDYAEVYLTVMGDG
 KGTGLGDCMWSVGVYAYEMIYGRSPFAGTSARTFNINMFORLTKPDPPKVSDF
 LDILOSLGOKERLKEFGLCOHPFSKIDMNNIRNSPPPPVPTLKSDDDSNPRE
 KNSWSSPCQJSPSGSEBELPVQFSKSLGILGSESVSGLDPAKTSMEKK
 LLIKSLDSDQKCHRMQENTRLRRSEVAVLSQEKVLEKASRYQSLLEDDLA
 TYITCESSILKSLLEQARMEVSEDDPAQLLHIREQSKLOELKEQYQVEMRL
 MNQLEEDVLSARRSDIYSELRSLAEFRKATSCQKILKADQKPEVGEY
 AKLEKINAOQKIQLEKLEKAVASTATELQNTROAKERALEKQKLEONDES
 SEGRKLVLEAELEKRAVOASQHLEVHLKQKQHEEIKVLNDQIKKDLAKS
 TLENNQREHEEHEKIKLSEQKAMINMDSKIRSLORIVELSEANKILANSILPT
 QNRKVAQEMISEIQOKRYLETQAGKPAQRKKEBQLEKISHDSDKRLIELET
 RLREVSLEHBEQLEKRLTEIQSLQRESQITLQARALESSQLQKTELEET
 TARESEIQITAHREIQKFDALNRSCTVITLDEQLNLTEDNALNQNPFYLSK

BASE COUNT 2277 a 2204 c 2220 g 1875 t
 ORIGIN

Query Match 94.8%; Score 5843.2; DB 9; Length 8576;
 Best Local Similarity 97.2%; Pred. No. 0;
 Matches 6036; Conservative 0; Mismatches 3; Indels 171; Gaps 2;

QY 1 ATGTTGAAGTTCAAATATGAGAGCGCGGAATCTTTGAGTCTGCTGCTGAGACCAT 60
 DB 54 ATGTTGAAGTTCAAATATGAGAGCGCGGAATCTTTGAGTCTGCTGCTGAGACCAT 113
 QY 61 GCCAGCGGAGCTCCAGGCTGAAATCTGTTCTCAGGAGGAAACACCTTTATGACTCAA 120
 DB 114 GCCAGCGGAGCTCCAGGCTGAAATCTGTTCTCAGGAGGAAACACCTTTATGACTCAA 113
 QY 121 CAGCAGATGCTCTCTCTTTCCGAGAAAGGATATAGATCCCTTTGTTCTTTGAA 180
 DB 174 CAGCAGATGCTCTCTCTTTCCGAGAAAGGATATAGATCCCTTTGTTCTTTGAA 233
 QY 181 GAATGAGTACAGCTGCTGATGAGATTAACAGCTGAGCAATTTGTCGGAAGTAT 240
 DB 174 CAGCAGATGCTCTCTCTTTCCGAGAAAGGATATAGATCCCTTTGTTCTTTGAA 233
 QY 234 GAATGAGTACAGCTGCTGATGAGATTAACAGCTGAGCAATTTGTCGGAAGTAT 293
 QY 241 TCCGACACCAATAGCTAGTTACAGAGAGCTTCAGCTTCGCAAAAGACCTTGAAGTACA 300
 DB 294 TCCGACACCAATAGCTAGTTACAGAGAGCTTCAGCTTCGCAAAAGACCTTGAAGTACA 353
 QY 301 AGCTTGAGGTTGAGTCACTTCTGAGAGTCAAGTGTGTAAGAGAAAGACACCGG 360
 DB 354 AGCTTGAGGTTGAGTCACTTCTGAGAGTCAAGTGTGTAAGAGAAAGACACCGG 413
 QY 361 GACATCTATGCTAAGAAAGTATGAGAGAAAGAGCTTTATGAGCCGAGAGAGTTC 420
 DB 414 GACATCTATGCTAAGAAAGTATGAGAGAAAGAGCTTTATGAGCCGAGAGAGTTC 473
 QY 421 TTTTGAAGAAAGCGGAAACATATATCTTCGAAAGCAAGCCCGTGAATCCCCCAATTA 480
 DB 474 TTTTGAAGAAAGCGGAAACATATATCTTCGAAAGCAAGCCCGTGAATCCCCCAATTA 533
 QY 481 CAGTATGCTTTGAGCAAAATATCACTTTATCTGCTGATGAGATATAGCTTGAAGG 540
 DB 534 CAGTATGCTTTGAGCAAAATATCACTTTATCTGCTGATGAGATATAGCTTGAAGG 593
 QY 541 GACTTGCTGTCATCTTTGAAATAGATATGAGAGCAAGTATGAGTAAATCTGATACGTTT 600
 DB 594 GACTTGCTGTCATCTTTGAAATAGATATGAGAGCAAGTATGAGTAAATCTGATACGTTT 653
 QY 601 TACCTAGCTAGCTGATTTTGGCTGTTCAAGAGCTTCATGTAGGATACGCTGATACGA 660
 DB 654 TACCTAGCTAGCTGATTTTGGCTGTTCAAGAGCTTCATGTAGGATACGCTGATACGA 713
 QY 661 GACATCAAGCTTGAAGACATTTCTGTTGACCGACAGAGACATCAAGCTGTGTGATTTT 720
 DB 714 GACATCAAGCTTGAAGACATTTCTGTTGACCGACAGAGACATCAAGCTGTGTGATTTT 773

QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCGTTTCATCTGATGGGATACGTGATGCA 660
 Db 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCGTTTCATCTGATGGGATACGTGATGCA 660
 QY 661 GACATCAAGCCTGAGAACATTTCTGTTGACCGACAGACACATCAAGCTGATGATTTT 720
 Db 661 GACATCAAGCCTGAGAACATTTCTGTTGACCGACAGACACATCAAGCTGATGATTTT 720
 QY 721 GGATCTGCGCGGAAAAATGAAATTCAAAAGATGATGATGCTCAAACTCCGATTTGGAC 780
 Db 721 GGATCTGCGCGGAAAAATGAAATTCAAAAGATGATGATGCTCAAACTCCGATTTGGAC 780
 QY 781 CCAGATTACATGAGCTCTGAGAGTGTGATGATGAAAGGAGATGAAAGGACCTAC 840
 Db 781 CCAGATTACATGAGCTCTGAGAGTGTGATGATGAAAGGAGATGAAAGGACCTAC 840
 QY 841 GGCCTTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 841 GGCCTTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 901 TCCCTCTGCGAGAGGAACTCTGCGAGAACTTCAATTAATTAATTAATTAATTAATTAAT 960
 Db 901 TCCCTCTGCGAGAGGAACTCTGCGAGAACTTCAATTAATTAATTAATTAATTAATTAAT 960
 QY 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020
 Db 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020
 QY 1021 TTGTTGCGCGCGAG 1080
 Db 1021 TTGTTGCGCGCGAG 1080
 QY 1081 TCTTAAATTTGATGAGAACCAATTCGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
 Db 1081 TCTTAAATTTGATGAGAACCAATTCGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
 QY 1141 TCTGAGATGACCTCTCAATTTTGAATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 Db 1141 TCTGAGATGACCTCTCAATTTTGAATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1201 CCGTGCAGCTGAGCCCTCAAGCTCTCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1260
 Db 1201 CCGTGCAGCTGAGCCCTCAAGCTCTCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1261 TACAGAGAGCACTGGGAGATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1320
 Db 1261 TACAGAGAGCACTGGGAGATTTCTGATGATGATGATGATGATGATGATGATGATGAT 1320
 QY 1321 CCGTGCAGCTGAGCCCTCAAGCTCTCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1380
 Db 1321 CCGTGCAGCTGAGCCCTCAAGCTCTCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1381 TCTCAGAGCACTGATGACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 Db 1381 TCTCAGAGCACTGATGACCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1441 GAGTGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 Db 1441 GAGTGAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 QY 1501 TCCCTCTGCGAG 1560
 Db 1501 TCCCTCTGCGAG 1560
 QY 1561 TTGAG 1620
 Db 1561 TTGAG 1620
 QY 1621 GATATCAG 1680
 Db 1621 GATATCAG 1680
 QY 1681 GTGAG 1740

Db 1681 GTGAG 1740
 QY 1741 CGAGATGATCTTACGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 Db 1741 CGAGATGATCTTACGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 QY 1801 CGAG 1860
 Db 1801 CGAG 1860
 QY 1861 GTGAG 1920
 Db 1861 GTGAG 1920
 QY 1921 CTCCAG 1980
 Db 1921 CTCCAG 1980
 QY 1981 AATATCCGCGAG 2040
 Db 1981 AATATCCGCGAG 2040
 QY 2041 GATTCCTTCTGAG 2100
 Db 2041 GATTCCTTCTGAG 2100
 QY 2101 GAG 2160
 Db 2101 GAG 2160
 QY 2161 ATCCAG 2220
 Db 2161 ATCCAG 2220
 QY 2221 AAACATGCGAG 2280
 Db 2221 AAACATGCGAG 2280
 QY 2281 CAGCAGATGAG 2340
 Db 2281 CAGCAGATGAG 2340
 QY 2341 AAG 2400
 Db 2341 AAG 2400
 QY 2401 ATTCTCAG 2460
 Db 2401 ATTCTCAG 2460
 QY 2461 CAG 2520
 Db 2461 CAG 2520
 QY 2521 CAAAG 2580
 Db 2521 CAAAG 2580
 QY 2581 CTGAG 2640
 Db 2581 CTGAG 2640
 QY 2641 AAGATCAG 2700
 Db 2641 AAGATCAG 2700
 QY 2701 CGGAG 2760
 Db 2701 CGGAG 2760
 QY 2761 CTACAGCTCTCTCTGAG 2820

Db 2761 CTAGAGCTCTCCCTGAGAGAGCGGAGTCAAGTTGACAGCCCTGAGAGGCTGACCGGCG 2820
Qy 2821 GCCCTGAGAGCCAGCTTCCGCAAGCGGAAACAAGAGCTGAGAGAGCAACAGAGAGCT 2880
Db 2821 GCCCTGAGAGCCAGCTTCCGCAAGCGGAAACAAGAGCTGAGAGAGCAACAGAGAGCT 2880
Qy 2881 GAAGAGAGAGATCCAGGAGCTCAGGCGCATAGAGATGAATTCAGAGCGAAATTTGATGCT 2940
Db 2881 GAAGAGAGAGATCCAGGAGCTCAGGCGCATAGAGATGAATTCAGAGCGAAATTTGATGCT 2940
Qy 2941 CTTGTAACAGCTGTAAGTATGTAATCAAGACCTGAGAGAGAGAGCTTAACAGCTGAGAG 3000
Db 2941 CTTGTAACAGCTGTAAGTATGTAATCAAGACCTGAGAGAGAGAGCTTAACAGCTGAGAG 3000
Qy 3001 GAGACGCTGAGATCCAGCAACCAAACTTCTAATTGCGCAAAATTCAGATGAGGCTTCT 3060
Db 3001 GAGACGCTGAGATCCAGCAACCAAACTTCTAATTGCGCAAAATTCAGATGAGGCTTCT 3060
Qy 3061 GGCAGCAACGAGATTTGTAACAATGCGAAGTGAAGTGAAGCCATCTCCGCGGAGATC 3120
Db 3061 GGCAGCAACGAGATTTGTAACAATGCGAAGTGAAGTGAAGCCATCTCCGCGGAGATC 3120
Qy 3121 ACGGAACGAGAGATGAGCTTACAGCGCAAGAAACGATGAGAGCTCTGAAGACAG 3180
Db 3121 ACGGAACGAGAGATGAGCTTACAGCGCAAGAAACGATGAGAGCTCTGAAGACAG 3180
Qy 3181 TGCAACATGCTGAGAGAAACAGATCATGATTTGAGAGCCCTAAACGATGAGCTGAGAA 3240
Db 3181 TGCAACATGCTGAGAGAAACAGATCATGATTTGAGAGCCCTAAACGATGAGCTGAGAA 3240
Qy 3241 AAAAGAGGAG 3300
Db 3241 AAAAGAGGAG 3300
Qy 3301 TGTCGAGTTGAG 3360
Db 3301 TGTCGAGTTGAG 3360
Qy 3361 GATGAGAGAGATCAACGAGTCTCCGCAAGTGTGAGAGTGAAGAGAGAGAGAGAGAG 3420
Db 3361 GATGAGAGAGATCAACGAGTCTCCGCAAGTGTGAGAGTGAAGAGAGAGAGAGAGAG 3420
Qy 3421 GAGATTCGCTGAG 3480
Db 3421 GAGATTCGCTGAG 3480
Qy 3481 GACAGCTCAATGAATGAG 3540
Db 3481 GACAGCTCAATGAATGAG 3540
Qy 3541 CAG 3600
Db 3541 CAG 3600
Qy 3601 TTACAG 3660
Db 3601 TTACAG 3660
Qy 3661 GAACTGAG 3720
Db 3661 GAACTGAG 3720
Qy 3721 GAAAGATTCAGAGTTCTCTATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Db 3721 GAAAGATTCAGAGTTCTCTATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Qy 3781 CAAGCAAGATTCAGTTGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Db 3781 CAAGCAAGATTCAGTTGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Qy 3841 GTTCTCTGAGAGATCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 3841 GTTCTCTGAGAGATCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900

Qy 3901 GAGCTAG 3960
Db 3901 GAGCTAG 3960
Qy 3961 GCCCAGCGAG 4020
Db 3961 GCCCAGCGAG 4020
Qy 4021 ATGCGCATGTCGAGATGAG 4080
Db 4021 ATGCGCATGTCGAGATGAG 4080
Qy 4081 GCGCGCATGTCGAG 4140
Db 4081 GCGCGCATGTCGAG 4140
Qy 4141 AAG 4200
Db 4141 AAG 4200
Qy 4201 ACAAAGT 4260
Db 4201 ACAAAGT 4260
Qy 4261 GAATGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Db 4261 GAATGTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
Qy 4321 CCTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4380
Db 4321 CCTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4380
Qy 4381 GGTCTCCAG 4440
Db 4381 GGTCTCCAG 4440
Qy 4441 AGAATTAACAAG 4500
Db 4441 AGAATTAACAAG 4500
Qy 4501 AAAGT 4560
Db 4501 AAAGT 4560
Qy 4561 GAGCTGTGCTTCCGAG 4620
Db 4561 GAGCTGTGCTTCCGAG 4620
Qy 4621 GCAATACAG 4680
Db 4621 GCAATACAG 4680
Qy 4681 ACTGTGTGAG 4740
Db 4681 ACTGTGTGAG 4740
Qy 4741 TGGGTGACCGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800
Db 4741 TGGGTGACCGGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800
Qy 4801 GGTGATGCTAACTGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 4860
Db 4801 GGTGATGCTAACTGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 4860
Qy 4861 ATGAAATGAG 4920
Db 4861 ATGAAATGAG 4920
Qy 4921 TAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4980
Db 4921 TAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4980

D 2953 GACACGCTGAACCTCAACCAAACTTCTACTTGTCCAAACAACCTCATGAGGCTTCT 3012
Q 3061 GGGCCCAACGACGAGATTGTACAACTGCGAAAGTGAAGTGAACAATCTCGCGCGGAGATC 3120
D 3013 GGGCCCAACGACGAGATTGTACAACTGCGAAAGTGAAGTGAACAATCTCGCGCGGAGATC 3072
Q 3121 ACGGAACGAGAGATGACGCTTACACGACGAGAGCAAGCAAGTGAAGGCTCTGAAAGACCAAG 3180
D 3073 ACGGAACGAGAGATGACGCTTACACGACGAGAGCAAGCAAGTGAAGGCTCTGAAAGACCAAG 3132
Q 3181 TGCAACAATGCTGAGAGAAACAGGTCATGATTTTGGAGGCTTAAACGATGAGCTGTAGAA 3240
D 3133 TGCAACAATGCTGAGAGAAACAGGTCATGATTTTGGAGGCTTAAACGATGAGCTGTAGAA 3192
Q 3241 AAAAGAGGAGCTGAGAGGCTGAGAGAGGCTCTGAGTGAAGAAATCCGAGTTTGGAG 3300
D 3193 AAAAGAGGAGCTGAGAGGCTGAGAGAGGCTCTGAGTGAAGAAATCCGAGTTTGGAG 3252
Q 3301 TGTGGGTTGAGAGCTGACAGAGATGCTGACACCGAGAAACAGACAGAGCGAGAGCC 3360
D 3253 TGTGGGTTGAGAGCTGACAGAGATGCTGACACCGAGAAACAGACAGAGCGAGAGCC 3312
Q 3361 GATAGCGGATCACCGAGCTCTCGCAGAGTGTGAGTGTGAGTGAAGAGCAAGGCT 3420
D 3313 GATAGCGGATCACCGAGCTCTCGCAGAGTGTGAGTGTGAGTGAAGAGCAAGGCT 3372
Q 3421 GAGATTCGCTGCTGACAGAGGCTCTCAAGAGCAAGGCTGAGAGCGAGAGCTCTCT 3480
D 3373 GAGATTCGCTGCTGACAGAGGCTCTCAAGAGCAAGGCTGAGAGCGAGAGCTCTCT 3432
Q 3481 GACAACTCAATGACTGCTGAGAGAGAGCAATGCTTGAATGAAATGCTCCGAGCTTA 3540
D 3433 GACAACTCAATGACTGCTGAGAGAGAGCAATGCTTGAATGAAATGCTCCGAGCTTA 3492
Q 3541 CAGCAGAGCTGAGAGCTGAGAGAGAGCTCAACAGAGGCTTCTGAAAGAGCAAGCCAA 3600
D 3493 CAGCAGAGCTGAGAGCTGAGAGAGAGCTCAACAGAGGCTTCTGAAAGAGCAAGCCAA 3552
Q 3601 TTACAGAGCAGATGAGACTGACAGAAATTCATTTTCGTGACTCAAGAGCTGCA 3660
D 3553 TTACAGAGCAGATGAGACTGACAGAAATTCATTTTCGTGACTCAAGAGCTGCA 3612
Q 3661 GAACCTCAATGAGGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
D 3613 GAACCTCAATGAGGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3672
Q 3721 GAACCAATTCAGGTTCTCTAATTCATGAAAGAGTGAAGAGCACTAATTCCTCA 3780
D 3673 GAACCAATTCAGGTTCTCTAATTCATGAAAGAGTGAAGAGCACTAATTCCTCA 3732
Q 3781 CAACCAATTCATGATTTTCTGCAAGCCAAATGAGCACTGCTTAAAGAAAG-- 3838
D 3733 CAACCAATTCATGATTTTCTGCAAGCCAAATGAGCACTGCTTAAAGAAAG 3792
Q 3839 -----AGTTCTCTGAGTAC 3855
D 3793 GGTATTATTAGTGAAG 3852
Q 3856 AATAGCTGAAGCTGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3915
D 3853 AATAGCTGAAGCTGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3912
Q 3916 CTTGAG 3975
D 3913 CTTGAG 3972
Q 3976 ACGAGCAACCAACCAATCAAG 4035
D 3973 ACGAGCAACCAACCAATCAAG 4032
Q 4036 ATGCTGAGTGAAG 4095
D 4033 ATGCTGAGTGAAG 4092

Q 4096 CGCAAGAAAGAGTCTTCAATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4155
D 4093 CGCAAGAAAGAGTCTTCAATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4152
Q 4156 CACAATATTCCTCAACGATTCACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4215
D 4153 CACAATATTCCTCAACGATTCACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4212
Q 4216 TGTCTGATACCGTCACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 4275
D 4213 TGTCTGATACCGTCACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 4272
Q 4276 TGTCAACCCAG 4335
D 4273 TGTCAACCCAG 4332
Q 4336 ACACATTCACCGAG 4395
D 4333 ACACATTCACCGAG 4392
Q 4396 GAGCCCAAG 4455
D 4393 GAGCCCAAG 4452
Q 4456 GAGCAG 4515
D 4453 GAGCAG 4512
Q 4516 GACAAATGAAG 4575
D 4513 GACAAATGAAG 4572
Q 4576 GAGGAG 4635
D 4573 GAGGAG 4632
Q 4636 GAGAGATTCATACATGAG 4695
D 4633 GCA----- 4635
Q 4696 AGAACCCTCACTTCTGAG 4755
D 4636 ----- 4635
Q 4756 GAATCAGTTGCGAG 4815
D 4636 -----GAG 4662
Q 4816 CTTGAG 4875
D 4663 CTTGAG 4722
Q 4876 CCCTTCAGTGAACAG 4935
D 4723 CCCTTCAGTGAACAG 4782
Q 4936 TTGAG 4995
D 4783 TTGAG 4842
Q 4996 AAG 5055
D 4843 AAG 4902
Q 5056 GTGAAG 5115
D 4903 GTGAAG 4962
Q 5116 CCAACATTTTGAAG 5175
D 4963 CCAACATTTTGAAG 5022

Qy	5176	GGGCTCGCATCTGTGCGAGCATGCGCCAGCAAAGTCGATCTTCGCGCTACAAAGAAAC	5235
Db	5023	GGGCTCTGCGACTCTGTGCGAGCCATGCCAGCAAAGTCGATCTTCGCGCTACAAAGAAAC	5082
Qy	5236	CTCAGCAAAATCTGCATCTCGGAAAGAGATGAGACTTCAGAGCCCTGCACTGTATCCAC	5295
Db	5083	CTCAGCAAAATCTGCATCTCGGAAAGAGATGAGACTTCAGAGCCCTGCACTGTATCCAC	5142
Qy	5296	TTGACCAATTACAGTATCTCATTGGAAACCAATAAATCTACAGAAATGACATGAGCAG	5355
Db	5143	TTGACCAATTACAGTATCTCATTGGAAACCAATAAATCTACAGAAATGACATGAGCAG	5202
Qy	5356	TACACGCTCGAGGAATTCCTGATPAAGATGACCAATTCCTTGGCACTGCTGTATTGCC	5415
Db	5203	TACACGCTCGAGGAATTCCTGATPAAGATGACCAATTCCTTGGCACTGCTGTATTGCC	5262
Qy	5416	GCCCTCTTCCACAGCTTCCCTGTCGCATCGTGGAGGGAGAACAGCGCAGGGCAGCAGAG	5475
Db	5263	GCCCTCTTCCACAGCTTCCCTGTCGCATCGTGGAGGGAGAACAGCGCAGGGCAGCAGAG	5322
Qy	5476	GAGTACTTGTCTGTGTTTCCACGAATTTTGGATGTTGTCGATTTCTTAACGAGAAGTAC	5535
Db	5323	GAGTACTTGTCTGTGTTTCCACGAATTTTGGATGTTGTCGATTTCTTAACGAGAAGTAC	5382
Qy	5536	CGCACACACGATCTCAAGTGGAGTCCGTACCTTTGGCTTTGGCTACAGAGAACCTAT	5595
Db	5383	CGCACACACGATCTCAAGTGGAGTCCGTACCTTTGGCTTTGGCTACAGAGAACCTAT	5442
Qy	5596	CTGTTTGTGACCCACTTCACTCACTCGAAGTAAATGAGATCCAGGCAAGCTCTCTCAGCA	5655
Db	5443	CTGTTTGTGACCCACTTCACTCACTCGAAGTAAATGAGATCCAGGCAAGCTCTCTCAGCA	5502
Qy	5656	GGGACCCCTCGCCGAGGAGTACTGTGACAATCCCGAAACCCGCGCTACCTGGGCGCTGCAAT	5715
Db	5503	GGGACCCCTCGCCGAGGAGTACTGTGACAATCCCGAAACCCGCGCTACCTGGGCGCTGCAAT	5562
Qy	5716	TCTCTCAGAGCGATTTATCTTGGCGCTCTCATACGACGATPAATTAAGGGTCATTTGCTGC	5775
Db	5563	TCTCTCAGAGCGATTTATCTTGGCGCTCTCATACGAGATPAATTAAGGGTCATTTGCTGC	5622
Qy	5776	AAAGGAAACCTCGTGAAAGGATCCGGGCACTGAACACCAACCGGGGCGCTCCACTTCCCGC	5835
Db	5623	AAAGGAAACCTCGTGAAAGGATCCGGGCACTGAACACCAACCGGGGCGCTCCACTTCCCGC	5682
Qy	5836	AGCAGCCCCCAACAGCGAGGCGCCACCCACGTACAAAGAGCAATCACCAACAGCGGTGCC	5895
Db	5683	AGCAGCCCCCAACAGCGAGGCGCCACCCACGTACAAAGAGCAATCACCAACAGCGGTGCC	5742
Qy	5896	TCCAGCCCGAGGCGCGCCCGGAGAGGCGCCGACCCCGAGAGGCCAAGAGCACCCCAACGC	5955
Db	5743	TCCAGCCCGAGGCGCGCCCGGAGAGGCGCCGACCCCGAGAGGCCAAGAGCACCCCAACGC	5802
Qy	5956	TACCGCAGAGGGCGGAGCCGAGCTCGCAGGAGCAAGTCTCTTGGCCGCGCCCTTGGAGCGA	6015
Db	5803	TACCGCAGAGGGCGGAGCCGAGCTCGCAGGAGCAAGTCTCTTGGCCGCGCCCTTGGAGCGA	5862
Qy	6016	GAGAAATCTCCCGGCGCGATTACTAGCACGCGGAGAGAGGCGGTCTCCCGCAGAGCTGTTT	6075
Db	5863	GAGAAATCTCCCGGCGCGATTACTAGCACGCGGAGAGAGGCGGTCTCCCGCAGAGCTGTTT	5922
Qy	6076	GAAGACAGCACAGAGGAGCGCGCTCTGCGGAGAGCCGAGAGAACCCCGCTTCCAGAGTG	6135
Db	5923	GAAGACAGCACAGAGGAGCGCGCTCTGCGGAGAGCCGAGAGAACCCCGCTTCCAGAGTG	5982
Qy	6136	AAACAAGTCTTGGGACCAAGTCTTTC 6158	
Db	5983	AAACAAGGAAAGGGCAGAGTGC 6005	

QY	721	GSATCTGCGGGAAAAATGAATTTCAAAACAAGATGGATGGCAAACTCCGATTGGGAC	780
Db	739	GGATCTGCGGGAAAAATGAATTTCAAAACAAGATGGATGGCAAACTCCGATTGGGAC	798
QY	781	CCAGATTACATGCTCTCTGAAGTGTGACTGTATGAAACGGGGATGGAAAAAGCACTTAC	840
Db	799	CCAGATTACATGCTCTCTGAAGTGTGACTGTATGAAACGGGGATGGAAAAAGCACTTAC	858
QY	841	GGCGTGGACTGTGACTGGTGCAGTGGGCGGTGATGGCTTATGAAATGATTTATGGGGA	900
Db	859	GGCGTGGACTGTGACTGGTGCAGTGGGCGGTGATGGCTTATGAAATGATTTATGGGGA	918
QY	901	TCCCGCTTGGCAGAGGGAACTCTGCGCAAGAACCTTCAATTAATATGAAATTTCCAGCGG	960
Db	919	TCCCGCTTGGCAGAGGGAACTCTGCGCAAGAACCTTCAATTAATATGAAATTTCCAGCGG	978
QY	961	TTTTTGAATTTCCAGATGACCCCAAGTGAAGTGAACCTTCTTGTGATCTGATTTCAAGC	1020
Db	979	TTTTTGAATTTCCAGATGACCCCAAGTGAAGTGAACCTTCTTGTGATCTGATTTCAAGC	1038
QY	1021	TTGTGTGCGGGCAGAAAAGAGACTGAAGTTTGAAGTCTTGTGCTACCTCTTCTC	1080
Db	1039	TTGTGTGCGGGCAGAAAAGAGACTGAAGTTTGAAGTCTTGTGCTACCTCTTCTC	1098
QY	1081	TCATAAATTGACTGGACAACATTGTAACTCTCTCCGCCCTTGGTCCCACTCTCAAG	1140
Db	1099	TCATAAATTGACTGGACAACATTGTAACTCTCTCTCCGCCCTTGGTCCCACTCTCAAG	1158
QY	1141	TCTGACGATGACCTCCCAATTTTGATGAACAGAGAAATGTGGGGTTTCACTCCT	1200
Db	1159	TCTGACGATGACCTCCCAATTTTGATGAACAGAGAAATGTGGGGTTTCACTCCT	1218
QY	1201	CGGTGCCAGCTGAGCCCTCTCAGGCTCTCTGGGTGAAGACTCGCTTGTGTGGGGTTTTCG	1260
Db	1219	CGGTGCCAGCTGAGCCCTCTCAGGCTCTCTGGGTGAAGACTCGCTTGTGTGGGGTTTTCG	1278
QY	1261	TACAGCAAGGCACTGGGGATTTCTTGATGATCTGAGTCTGTTGTGTGGGTTCTGACCTCC	1320
Db	1279	TACAGCAAGGCACTGGGGATTTCTTGATGATCTGAGTCTGTTGTGTGGGTTCTGACCTCC	1338
QY	1321	CCTGCGAAGACTGCTCCATGGAAAGAAACCTTCTATCAAAAGCAAGAGCTACAAAC	1380
Db	1339	CCTGCGAAGACTGCTCCATGGAAAGAAACCTTCTATCAAAAGCAAGAGCTACAAAC	1398
QY	1381	TCTCAGACAAAGTGTCAACAGATGAGACAGAAATGACCCGGTTACATCGAGAGTGTCA	1440
Db	1399	TCTCAGACAAAGTGTCAACAGATGAGAGCAAGAAATGACCCGGTTACATCGAGAGTGTCA	1458
QY	1441	GAGGTGGAGGCTGTGCTTATGTACAGAAAGAGGTGAGGCTGAAAGGCTCTGAGACTCAGGA	1500
Db	1459	GAGGTGGAGGCTGTGCTTATGTACAGAAAGAGGTGAGGCTGAAAGGCTCTGAGACTCAGGA	1518
QY	1501	TCCCTCTGGAGCAGAGCCTTGCTACCTAACACAGAGATGCAATGACTTAAAGCAAGT	1560
Db	1519	TCCCTCTGGAGCAGAGCCTTGCTACCTAACACAGAGATGCAATGACTTAAAGCAAGT	1578
QY	1561	TTGGAGCAAGACAGATGAGAGTGTCCACAGAGATGACAAAGACTGTGACCTTCCAT	1620
Db	1579	TTGGAGCAAGACAGATGAGAGTGTCCACAGAGATGACAAAGACTGTGACCTTCCAT	1638
QY	1621	GATATCAGAGAGCAGAGCCGGAGGCTCCAGAAATCAAAAGACAGAGATACAGAGCTCAA	1680
Db	1639	GATATCAGAGAGCAGAGCCGGAGGCTCCAGAAATCAAAAGACAGAGATACAGAGCTCAA	1698
QY	1681	GTGGAAAGAAATGAGGTTGATGATGAATCAGTTGGAAGGAACTTGTCTCAGCAAGAGA	1740
Db	1699	GTGGAAAGAAATGAGGTTGATGATGAATCAGTTGGAAGGAACTTGTCTCAGCAAGAGA	1758
QY	1741	CGAGATGATCTCTAGATCTGAGGCTGAGAGAGTCTCGGCTTGTGCTGGAAGATTTCAAG	1800
Db	1759	CGAGATGATCTCTAGATCTGAGGCTGAGAGAGTCTCGGCTTGTGCTGGAAGATTTCAAG	1818
QY	1801	CGAAAGCGACGAATGTGACGATTAACCTGTTAAAGGCTTAAGATCAAAGGAAGCTGAA	1860

[illegible]

OY	5056	GTGAAGAAAGTAAACATGCTCCCTGGCCAGTCCCACTGCTGGCCAGCCGACATCTCA	5115
Db	4921	GTGAAGAAAGTAAACATGCTCCCTGGCCAGTCCCACTGCTGGCCAGCCGACATCTCA	4980
OY	5116	CCCAACATTTTGAAGCTGTCAAGGCGTCCCATCTTTTGGGCGAGGCAAGTTGAGAC	5175
Db	4981	CCCAACATTTTGAAGCTGTCAAGGCGTCCCATCTTTTGGGCGAGGCAAGTTGAGAC	5040
OY	5176	GGGCTCTGCATCTGTGCGAGCATGCGCCAGCAAAAGTCGTCAATCTCCGCTACACGAAAC	5235
Db	5041	GGGCTCTGCATCTGTGCGAGCATGCGCCAGCAAAAGTCGTCAATCTCCGCTACACGAAAC	5100
OY	5226	CTCAGCAATATCTGCATCCGGAAGAGATAGAGACTCAGAGCCTTGAGCTGTATCCAC	5295
Db	5101	CTCAGCAATATCTGCATCCGGAAGAGATAGAGACTCAGAGCCTTGAGCTGTATCCAC	5160
OY	5296	TTCAACCAATTCAGATATCTCATTTGGAACCAATAAATTCAGAAATGCAATGAGAGAG	5355
Db	5161	TTCAACCAATTCAGATATCTCATTTGGAACCAATAAATTCAGAAATGCAATGAGAGAG	5220
OY	5336	TACACGCTCGAGGAATTCCTGGATPAAGATGCCATTCCTTGGCACCTGCTGTGTTGCC	5415
Db	5221	TACACGCTCGAGGAATTCCTGGATPAAGATGCCATTCCTTGGCACCTGCTGTGTTGCC	5280
OY	5416	GCTCTTTCCAACAGCTTCCTCTGTCTCAATCGTGCAGGTGAACAGCGCAGCGACGAGAG	5475
Db	5281	GCTCTTTCCAACAGCTTCCTCTGTCTCAATCGTGCAGGTGAACAGCGCAGCGACGAGAG	5340
OY	5476	GAGTACTTGTCTGTGTTTCCACGAAATTTGGAGTGTTCGTGATTTCTTACGGAAGCTATGC	5535
Db	5341	GAGTACTTGTCTGTGTTTCCACGAAATTTGGAGTGTTCGTGATTTCTTACGGAAGCTATGC	5400
OY	5536	CGCACAGACGATCTTCAGTGGAGTTCGCTTACCTTTGGCTTTGGCTTACAGAAACCTCAT	5595
Db	5401	CGCACAGACGATCTTCAGTGGAGTTCGCTTACCTTTGGCTTTGGCTTACAGAAACCTCAT	5460
OY	5596	CTGTTTGTGACCCACATTTCATCTCTGAAAGTAAATTTGAGATTCAGAGGACGCTCTCAGCA	5655
Db	5461	CTGTTTGTGACCCACATTTCATCTCTGAAAGTAAATTTGAGATTCAGAGGACGCTCTCAGCA	5520
OY	5656	GCGACCCCTTGGCCGAGCGTACTGTGACATCCCGAACCCGCGCTACCTTGAGGCGCTGCATT	5715
Db	5521	GCGACCCCTTGGCCGAGCGTACTGTGACATCCCGAACCCGCGCTACCTTGAGGCGCTGCATT	5580
OY	5716	TTCCTCAGAGAGGATTTACTTGGCGTCTCTCATPACAGAGATTAATTAAGGCTCATTTGCTGC	5775
Db	5581	TTCCTCAGAGAGGATTTACTTGGCGTCTCTCTCATPACAGAGATTAATTAAGGCTCATTTGCTGC	5640
OY	5776	AAGGAAACCTTCGTAGAGAGTTCGCGACCTGAAACAACAACCGGGGCCCTTCACCTCCGC	5835
Db	5641	AAGGAAACCTTCGTAGAGAGTTCGCGACCTGAAACAACAACCGGGGCCCTTCACCTCCGC	5700
OY	5836	AGCAGCCCCCAACAACGAGGCGCCACCCAGTACAAAGAGCATATCACCAAGCGGTGGCC	5895
Db	5701	AGCAGCCCCCAACAACGAGGCGCCACCCAGTACAAAGAGCATATCACCAAGCGGTGGCC	5760
OY	5896	TCCAGCCCAAGCGCGCCCGGAAAGGCCCCACGACCCCGCGAGAGCCAAAGACACCCCAACGC	5955
Db	5761	TCCAGCCCAAGCGCGCCCGGAAAGGCCCCACGACCCCGCGAGAGCCAAAGACACCCCAACGC	5820
OY	5956	TACCGCAGAGGGGCGGACCGAGGTGCGAGGGGCAAGTCTCCGCGCGCCCTCGTGAAGGA	6015
Db	5821	TACCGCAGAGGGGCGGACCGAGGTGCGAGGGGCAAGTCTCCGCGCGCCCTCGTGAAGGA	5880
OY	6016	GAGAAATCCCCCGCGCGGATATCTACAGACGCGGAGAGAGCGGTCTCCCGCAGAGCTGTTT	6075
Db	5881	GAGAAATCCCCCGCGCGGATATCTACAGACGCGGAGAGAGCGGTCTCCCGCAGAGCTGTTT	5940
OY	6076	GAAAGCACACAGAGGCGCGGTGCGCTGCGGGAAGCGGAGAGACCCCGCTATCCAGAGTG	6135
Db	5941	GAAAGCACACAGAGGCGCGGTGCGCTGCGGGAAGCGGAGAGACCCCGCTATCCAGAGTG	6000
OY	6136	AACAAAGTCTGGAACCACTTTC	6158

DB	6001	AACAAGGAGAGAGGCGACAGTGC	6023
RESULT 13	AF086824	6954 bp	mRNA linear ROD 11-NOV-1998
LOCUS	AF086824		
DEFINITION	Mus musculus rho/rac-interacting citron kinase (Crik) mRNA,		
ACCESSION	AF086824		
VERSION	AF086824.1	GI:3599508	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Di Cunto, F., Calautti, E., Hsiao, J., Ong, L., Topley, G., Turco, E. and Dotto, G.P.		
TITLE	Citron rho-interacting kinase, a novel tissue-specific ser/thr kinase encompassing the Rho-Rac-binding protein Citron		
JOURNAL	J. Biol. Chem. 273 (45), 29706-29711 (1998)		
MEDLINE	99009084		
PUBMED	9792683		
REFERENCE	2 (bases 1 to 6954)		
AUTHORS	Di Cunto, F., Calautti, E., Hsiao, J., Ong, L., Topley, G., Turco, E. and Dotto, G.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-AUG-1998) Biology, University of Torino, Via Santena 5 bis, Torino, TO 10126, Italy		
AUTHORS	Location/Qualifiers		
FEATURES	1..6954		
source	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/tissue_type="spleen"		
	1..6954		
	/gene="Crik"		
	772..6939		
	/gene="Crik"		
	/function="protein kinase"		
	/note="Crik; rho/rac binding protein; alternatively spliced form of citron encoded by GenBank Accession Number AF099218; similar to Mus musculus citron-K encoded by GenBank Accession Number AF070066; similar to Rattus norvegicus citron-K encoded by GenBank Accession Number AF070065"		
	/codon_start=1		
	/product="rho/rac-interacting citron kinase"		
	/protein_id="AAC72823.1"		
	/db_xref="GI:3599509"		
	/translation="MLKFKYGVNPSPASSEPIASPSRLNLPFGCKPPLMTQOQMSALSRGMLDLPALFEKSCSPALMKMKHVSFPQKYSDTTALRLDPSARDPFRSLVGGHAFNVDYVREKATGVDYFANKIKKKKLLQBOQVFFBEENLIIVASVSPITPOLQYAAQDNRLYLWVEYQGGDFSLNRYEDQDSNIQETPDYMAEVLTVNMEDGVVHRDLPENNLIIDRTGELKLVDPGSAATNSNKNYDAKPIGTDPMAEVLTVNMEDGVVGTVLDICDMMWSGVAVEMVYGTPTPTGEGSARTFNNINMFOHFLKFPDPKXVSEILLDIQSLILCVQKERLKEFGLCCHPFPATDMNNIRNSPPVPTLKSDDPNSPDEPKELNSMVAFLTPAEPPLAFSGEELPVGFSYSKALVIGBSVSVSLSDPANVSWEKLLLSKSELSDSDCKCHMCEOMTRLHPRVSEVETLQKVEVLEKASTPQBSLLRDLLNTYTFSSLSKSLSEARAEVQSEDPKALQLLHDIRESGRKQLTEKDYQKQVEMRLMNMQLEEDVARSRRSDIYESLEBSRLAAEFPRKAECHQKLMKAKDOQGEVGEYSKLEKINAEQQLIQLQELQEKLEKA VKASTETELLQNRQKERAERLEKLNHEDSSSGEGIKTKLVEAEERHSLLENKVKRLQETMERBNRLKODIQKSDIQMDAKLLEBKKHREYQSAQHLVHLKQEQHYEKKIYVLNDQIKDIALNKSLENNQARHEEAAHKKQKLSIQEQAAMINAMDSKIRSLDEORIVELSEANKLAANSGLFQRMKAQEMNISELRQOKFYLEVTDAGLEAKNORLEBOLERKISHQSDSRILEFRLREVLHBEHOKLEIKRLQLEHLOSLQERESQTLQCARALLESQLRQKTLRETTAAFEELQALTAHREDEIQKRPDLNRSCTVITLBEQLNQUTEDNAELNNQNYLSKQIDASGANDELYQASLEVDLRKEITREHQLTSQKQVTEBALKTTQMLBEQVQVLEALNDELKQKQZKARSVLDEGDSQFEREQLQMLDTEKQSAARDITRESROVLEAVEHKAETLALQOALKEQKAKAESLSDKLNLEKHAMENMARSQQKQKLETRKORLEQLQALQOQMDLQKNHIFLITQLOJLEALDRADILKTERSLEVEQLNENIQVLYSEKQVMEGTSIQOQK		

LIDELAKMDPAKKKKVPLQVNELEKLEKAKRABELEALQTRLEISAREEA
 HKRATHPHSTPAIRKQOLMSALVRSSEHOPMSLSLAPSSSRKSSSTPEERSR
 LKERHMHNIPIHFNVLNMRKATKCAVCLTVHGRQASCLCYWCHPKSCCLPAT
 CGLPAAYATHFEAFCDKXNPSBGLQSKBPSLHLEGMKPRANKKQOQMDVKYI
 VLEBSKVLVYDNEBARAGORPVEBELCJPDGVSJHGAVSELANTAKQVMDVKYI
 MESHPTTCWPGRTLYLAPSPDKORWTALSVYAGRPVREKADPAKLVNLSL
 KLEGRDLNCTLPSPDOVLVGTREGLVYALNLSNLTTHBGIAGPOLYIYDLE
 KLIMJEBRALCLDVKKVKSIALOSHIAPODPSNPIHEVKGCHLPAKIKNSL
 CTCAMPKSVILIRYNDLSKICREKITESESCGTHNTSILIGKVFIDMKY
 YLIDFLDKNDHSLAPVAPSSNSFPVSIQANSQOREYLLCFHGRVFDVGR
 RSRDULKMSRLFLAFAYREPLYLVTHFNSLEVLBIQANSISPARAYLIPNRYL
 GPAISSGAIYLAASSYODKLRLVLCCKGNLVKESGTEOHVPSYRSRSPMKRPLTNEH
 ITRVASSPAPRPSHPRPSCTPHRYRDEGRTELRDRSGRPLERBSKGRMLST
 RRRSPGRFEDSRRLPAGAYRPLPSQVKNVMDQSSV"

BASE COUNT 1779 a 1881 c 1972 g 1322 t
 ORIGIN

Query Match 82.8%; Score 5105.2; DB 10; Length 6954;
 Best Local Similarity 89.6%; Pred. No. 0;
 Matches 5533; Conservative 0; Mismatches 628; Indels 11; Gaps 4;

QY 1 ATGTGAAGTTCAAAATATGAGGCGGAATCCCTTGATGCTGCTGCTCAACCAT 60
 Db 772 ATGTGAAGTTCAAAATATGAGGCGGAATCCCTTGATGCTGCTGCTCAACCAT 831
 QY 61 GCCAGCGGCTCCAGGCTGAATCTGCTTCCAGGGGAAACACCTTTATGACTCA 120
 Db 832 GCCAGCTGGGCTCCAGGCTGAATCTGCTTCCAGGGGAAACACCTTATGACTCA 891
 QY 121 CAGCAGATGCTGCTCTTCCGGAAGGATATGATGCTGCTGCTGCTGCTGCTGCT 180
 Db 892 CAGCAGATGCTGCTCTTCCGGAAGGATATGATGCTGCTGCTGCTGCTGCTGCT 951
 QY 181 GAATGAGTGAAGCTGCTGATGAAGATTAAGACGAGCACTTTGCTCGGAAGTAT 240
 Db 952 GAATGAGTGAAGCTGCTGATGAAGATTAAGACGAGCACTTTGCTCGGAAGTAT 1011
 QY 241 TCCGACACCACTAGTGAAGTGAAGGCTCCAGCTTCCGGAAGGATTTGAGTCA 300
 Db 1012 TCCGACACCACTAGTGAAGTGAAGGCTCCAGCTTCCGGAAGGATTTGAGTCA 1071
 QY 301 AGCTTGTAGTGTGCTCACTTGTGCTGAAGTGAAGTGAAGGATTAAGGATTA 360
 Db 1072 AGCTTGTAGTGTGCTCACTTGTGCTGAAGTGAAGTGAAGGATTAAGGATTA 1131
 QY 361 GACATCTATCTATGAAGTGAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 420
 Db 1132 GACATCTATCTATGAAGTGAAGGATTAAGGATTAAGGATTAAGGATTAAGGATTA 1191
 QY 421 TTTTGAAGGAGGAGGAGATATATCTGAGGAGCAAGCCCGTGAATCCCAATTA 480
 Db 1192 TTTTGAAGGAGGAGGAGATATATCTGAGGAGCAAGCCCGTGAATCCCAATTA 1251
 QY 481 CAGTATGCTTTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
 Db 1252 CAGTATGCTTTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1311
 QY 541 GACTGCTGCTCACTTTGAATAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
 Db 1312 GACTGCTGCTCACTTTGAATAGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1371
 QY 601 TACCTAGCTGAGCTGATTTGCTGCTCAAGGCTTCACTGATGAGGATACGTGATCA 660
 Db 1372 TACCTAGCTGAGCTGATTTGCTGCTCAAGGCTTCACTGATGAGGATACGTGATCA 1431
 QY 661 GACATCAAGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
 Db 1432 GACATCAAGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
 QY 721 GATCTGCTGCGGAAATGAATGAATCAAGATGATGAATGCAACTCCGATTTGGAGCC 780
 Db 1492 GATCTGCTGCGGAAATGAATGAATGAATCAAGATGATGAATGCAACTCCGATTTGGAGCC 1548

QY 781 CCAATTACATGCTCTCTGAAGTCTGATCTGTATGAATGAGGAGTGGAAAAGGACCTAC 840
 Db 1549 CCAATTACATGCTCTCTGAAGTCTGATCTGTATGAATGAGGAGTGGAAAAGGACCTAC 1608
 QY 841 GGCCTGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 Db 1609 GGCCTGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
 QY 901 TCCCCCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 Db 1669 ACCCATTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1728
 QY 961 TTTTGAATTTCCAGATGAGCCCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 Db 1729 TTTTGAATTTCCAGATGAGCCCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1788
 QY 1021 TTGTTGTCGCGCGAGAAAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1080
 Db 1789 TTGTTGTCGCGCGAGAAAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1848
 QY 1081 TCTAAATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 Db 1849 GCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1908
 QY 1141 TCTGACGATGACACTTCAATTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1199
 Db 1909 TCTGACGATGACACTTCAATTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1968
 QY 1200 TCCGTGTCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1259
 Db 1969 TCCGTGTCAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2027
 QY 1260 GTACAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1319
 Db 2028 GTACAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2087
 QY 1320 CCGTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1379
 Db 2088 CCGTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2147
 QY 1380 CTCTCAGAGCAAGTGTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1439
 Db 2148 CTCTCAGAGCAAGTGTCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2207
 QY 1440 AGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1499
 Db 2208 AGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2267
 QY 1500 ATCCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1559
 Db 2268 ATCCCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2327
 QY 1560 TTTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1619
 Db 2328 TTTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2387
 QY 1620 TGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1679
 Db 2388 TGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2447
 QY 1680 AGTGAAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1739
 Db 2448 AGTGAAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2507
 QY 1740 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1799
 Db 2508 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2567
 QY 1800 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1859
 Db 2568 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2627
 QY 1860 AGTGAAGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1919

Db 2628 AATGGAGAAATATTCCTCAAACTGGAGAAAGTCAATGCTTGAGCAGACCTGAAAGTCCAGGA 2687
Qy 1920 GCTCCAAAGGAACTGGAGAAAGCTGTAAAGCCAGACAGAGGCCACCGAGCTGTGCA 1979
Db 2688 GCTCCAGAGAGAGCTGGAAGAGCTGTAAAGCCAGACAGAGGCCACCGAGCTGTGCA 2747
Qy 1980 GAAATATCCGCTCAGGCAAGAGAGCCAGAGGAGCTGTGAGAACTGTGAGAACCGAGA 2039
Db 2748 GAAATATCCGCTCAGGCAAGAGAGCCAGAGGAGCTGTGAGAACTGTGAGAACCGAGA 2807
Qy 2040 GATATCTTGAAGGATCAGAAAGAGCTGTGAGAGCTGTGAGAAAGCCGCTCATTTCTT 2099
Db 2808 AGATCTCTCCAGAGGATTAAGAGCTGTGAGAGCCGAGAGAAAGCCGCTCATTTCTT 2867
Qy 2100 GGAGAACAAAGTAAAGAGCTAGAGACATGAGAGGCTAGAGAAAAAGAGCTAGAGATGA 2159
Db 2868 GGAGAACAAAGTAAAGAGCTAGAGACATGAGAGGCTAGAGAACAGCTAGAGATGA 2927
Qy 2160 CATCCAGCAAAATCCCAACAGATCCAGAGATGCTGATTAATTTCTGAGCTGAGAA 2219
Db 2928 CATCCAGCAAAAGTCCGAACAGATCCAGAGATGCTGATTAATTTCTGAGCTGAGAA 2987
Qy 2220 GAAACATCGGAGGCCCAAGTCTCAGCCAGACCTAGAAAGTGCACCTGAGAACAGAGAA 2279
Db 2988 GAAACATCGGAGGCTCAGAGTCTCAGCTCAACATCTAGAGATCACTTGAAGAGAGGA 3047
Qy 2280 GCAGACTATGAGGAAAAAGATTAAAGTGTGACATAGATTAAGAAAGACCTGGCTGA 2339
Db 3048 ACAGACATACGAGAAAAAGATCAGATTAAGATCAATCAATTAAGAAAGACCTGGCTGA 3107
Qy 2340 CAAGGAGACACTGAGGAAATGATGACAGAGACAGAGAGAGAGGCCCATGAGAAAGGCAA 2399
Db 3108 CAAGGAGAGCTTGAGAAATGATGACAGAGACAGAGAGAGAGGCCCATGAGAAAGGCAA 3167
Qy 2400 AATTCTCAGCGAACAGAAAGCGATGATCATGCTATGATTCAGAGATCAGATCCCTGGA 2459
Db 3168 GATCCTCAGCGAGAGAAAGGAGATGATCAAGCGATGATTCAGAGATCCCTGGA 3227
Qy 2460 ACAGAGATTTGTGAATCTGTCTGAAGCCATTAATTTGAGCAATTAAGCAAGTCTTTTAC 2519
Db 3228 GCAGAGATCGTGGAGCTGTGAGAGCCAGAACAGCTTCGCGCAACAGACGCTCTTCA 3287
Qy 2520 CCAAGAGAACTGAAGGCCAGAGAGAGATGATTTCTGAATCTCAGAGCAAGAAATTTTA 2579
Db 3288 CCAAGAGAACTGAAGGCCAGAGAGATGATTTCTGAATCTCAGAGCAAGAAATTTTA 3347
Qy 2580 CCTGAGACACAGAGCTGGAGAAATTTGAGAGCCAGAACCGAAACTGAGAGAGCTGGA 2639
Db 3348 CCTGAGACACAGAGCCGAGAAAGCTGAGAGCCAGAACCGAAAGCTGAGAGAGCTGGA 3407
Qy 2640 GAAATACGCCAACCAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2699
Db 3408 GAAATACGCCAACCAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3467
Qy 2700 GCGAGAGTCAAGTCAAGAT 2759
Db 3468 GAGGAGTCAAGCTCTGAT 3527
Qy 2760 GCTAAGTCTCTCTGAT 2819
Db 3528 GCTAAGTCTCTCTGAT 3587
Qy 2820 GGCCCTGAT 2879
Db 3588 AGCCCTGAT 3647
Qy 2880 TGAAGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2939
Db 3648 GGAAGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3707
Qy 2940 TCTTCTGAACAGCTGTCTGTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2999

Db 3708 CCTTGCAACAGCTGACCGATCATCACAGCTTGAGAGAGAGAGAGAGAGAGAT 3767
Qy 3000 GGAACAAGCTGAATCAACCAACCAAACTTTCTTGTCCAAACCACTGAGAGAGCTTC 3059
Db 3768 GGAACAAGCTGAGCTCAACCAACCAAACTTTCTTGTCCAAACCACTGAGAGAGCTTC 3827
Qy 3060 TGGCGCCAAAGAGAGATTTGTCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAT 3119
Db 3828 CGGGGCAATGAGAGATTTGTCAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAT 3887
Qy 3120 CACGGAACAGAGATGACAGCTTACAGCCAGAGAACAGATGAGAGAGCTCTGAGAGAC 3179
Db 3888 CACGGAACAGAGATGACAGCTTACAGCCAGAGAACAGATGAGAGAGCTCTGAGAGAC 3947
Qy 3180 GTGCAACATGCTGAGAGAGAACAGATGATTTGAGAGAGAGAGAGAGAGAGAT 3239
Db 3948 ATGCAACATGCTGAGAGAGAACAGATGATTTGAGAGAGAGAGAGAGAGAGAT 4007
Qy 3240 AAAAGAGGCGCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3299
Db 4008 GAAAGAGGCGCAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4067
Qy 3300 GTTGGGTTTGAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3359
Db 4068 GTTGGGTTTGAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4127
Qy 3360 CGATCAGCGAGATCAACAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAT 3419
Db 4128 CGATCAGCGAGATCAACAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAT 4187
Qy 3420 TGAATTTCTGCTTGCAGAGAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAT 3479
Db 4188 CGATCAGCGAGATCAACAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAT 4247
Qy 3480 TGAAGCTCAATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3539
Db 4248 GAGCAAGCTCAACAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAT 4307
Qy 3540 ACAGAGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3599
Db 4308 ACAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4367
Qy 3600 ATTACAGAGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3659
Db 4368 ATTACAGAGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4427
Qy 3660 AGAAGCTTAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3719
Db 4428 GAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4487
Qy 3720 GGAAGAT 3779
Db 4488 GGAAGAT 4547
Qy 3780 ACAAACCAACTCATGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAT 3839
Db 4548 GCAAACCAACTCATGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAT 4607
Qy 3840 GATTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3899
Db 4608 GATTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4667
Qy 3900 AGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3959
Db 4668 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4727
Qy 3960 TGCCCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4019
Db 4728 TGCCCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4787
Qy 4020 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4079
Db 4788 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 4847

Db 961 CCGA----- 964
Qy 2138 GAGAAAAAGAGTGAAGATGACATCCAGCAAAAATCCCAACAGATCCAGCATGGCTG 2197
Db 965 ----- 964
Qy 2198 ATAAATTTGGAGCTCGAAGAGAAACATCGGAGGCCAAGTCTACGCCAGACCTAG 2257
Db 965 ----- 964
Qy 2258 AAGTGCACCTGAAAACAGAAAAGAGACATCTGTGAGAAAAGATTAAAGTTTGAACATC 2317
Db 1015 AAGTACCTTGAACAGAAAGAAACAGACCTACAGAAAAGATCAAGTATTGGAACATC 1074
Qy 2318 AGATAAAGAAAGACCTGCTGACAGAGACACCTGAGAACATGATGACAGACAGAG 2377
Db 1075 AGATAAAGAAAGACCTGCTGACAGAGAGACCTGAGAACATGATGACAGACAGAG 1134
Qy 2378 AGAGAGCCCATGAGAAAGGCAAAATTTCTAGCCGAAAGAGAGGATGATCATGCTATG 2437
Db 1135 AGAGAGCCCATGAGAAAGGCAAAATTTCTAGCCGAAAGAGAGGATGATCATGCTATG 1194
Qy 2438 ATTCCAGATCAGATCCCTGGAACAGAGATTTGGAACCTGTCTGAAGCCAAATTAATG 2497
Db 1195 ATTCCAGATCAGATCCCTGGAACAGAGATTTGGAACCTGTCTGAAGCCAAATTAATG 1254
Qy 2498 CAGCAATAGCAGCTTTTTCACCAAGAAAGACATGAAGGCCAAGAAAGATGATTTG 2557
Db 1255 CAGCAATAGCAGCTTTTTCACCAAGAAAGACATGAAGGCCAAGAAAGATGATTTG 1314
Qy 2558 AACTCAGGCAACGAAATTTTACCTGAGACACAGAGCTGGGAAAGTTGGAGGCCAAGAC 2617
Db 1315 AACTCAGGCAACGAAATTTTACCTGAGACACAGAGCTGGGAAAGTTGGAGGCCAAGAC 1374
Qy 2618 GAAATCTGAGAGACAGCTGAGAGATGACCAACCAAGACCAAGTGAACAAATCGGC 2677
Db 1375 GAAATCTGAGAGACAGCTGAGAGATGACCAACCAAGATGACCAAGTGAACAAATCGGC 1434
Qy 2678 TGCTGGAACCTGAGACCAAGATTTGCGGAGGTCACTCTAGAGCAAGAGAGCAAAACTGG 2737
Db 1435 TGCTGGAACCTGAGACCAAGATTTGCGGAGGTCACTCTAGAGCAAGAGAGCAAAACTGG 1494
Qy 2738 AGCTCAAGCCAGCTGACAGAGTACAGCTCTCCCTGACAGAGCGGAGTCAAGTTG 2797
Db 1495 AGCTCAAGCCAGCTGACAGAGTACAGCTCTCCCTGACAGAGCGGAGTCAAGTTG 1554
Qy 2798 CAGCCCTGCAAGCTGCAACGCGCGCTGAGAGCCAGCTTGCCTCAAGGCGAAGACAG 2857
Db 1555 CAGCCCTGCAAGCTGCAACGCGCGCTGAGAGCCAGCTTGCCTCAAGGCGAAGACAG 1614
Qy 2858 TGGAAGAGACCAAGAGAGAGTGAAGAGATCCAGAGCACTCAAGGCAATAGAGT 2917
Db 1615 TGGAAGAGACCAAGAGAGAGTGAAGAGATCCAGAGCACTCAAGGCAATAGAGT 1674
Qy 2918 AATTCAGCGCAATTTGATGCTTTGTAACAGCTGATCTGTATCAAGACCTGAG 2977
Db 1675 AATTCAGCGCAATTTGATGCTTTGTAACAGCTGATCTGTATCAAGACCTGAG 1734
Qy 2978 AGAGCTCAAGCTGACAGAGTCAAGAGTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 3037
Db 1735 AGAGCTCAAGCTGACAGAGTCAAGAGTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 1794
Qy 3038 CCAAACTGATGAGGCTTCTGAGCCCAAGAGAGATTTGATCAAGCTGAGAGTGA 3097
Db 1795 CCAAACTGATGAGGCTTCTGAGCCCAAGAGAGATTTGATCAAGCTGAGAGTGA 1854
Qy 3098 TGAACCTCTCCGCTGGAGATCAAGAGAGATGACGCTTACAGCCAGAGCAAA 3157
Db 1855 TGAACCTCTCCGCTGGAGATCAAGAGAGATGACGCTTACAGCCAGAGCAAA 1914
Qy 3158 CGATGAGGCTCTGAGAGACAGCTGACCAATGCTGAGAGCAAGCTCAAGTGAAG 3217
Db 1915 CGATGAGGCTCTGAGAGACAGCTGACCAATGCTGAGAGCAAGCTCTGAGAG 1974

Qy 3218 CCTAAACGATGAGCTGATGAAAAAGAGCGGAGAGTGGAGGCTTGAAGAGCGCTG 3277
Db 1975 CCTAAACGATGAGCTGATGAAAAAGAGCGGAGAGTGGAGGCTTGAAGAGCGCTG 2034
Qy 3278 GTGATGAGAAATCCAGTTTGAAGTCTGGGTTTGAAGAGCTGAGAGATGCTGACAC 3337
Db 2035 GTGATGAGAAATCCAGTTTGAAGTCTGGGTTTGAAGAGCTGAGAGATGCTGACAC 2094
Qy 3338 AGAAACAGAGCAGGCGGAGAGCTGACGCGATCAAGGCTTCCAGGCTGAGAGC 3397
Db 2095 AGAAACAGAGCAGGCGGAGAGCTGACGCGATCAAGGCTTCCAGGCTGAGAGC 2154
Qy 3398 TGGCAGTGAAGAGCAGAGCTGATCTGCTGAGAGAGCTGCAAGAGAGCA 3457
Db 2155 TGGCAGTGAAGAGCAGAGCTGATCTGCTGAGAGAGCTGCAAGAGAGCA 2214
Qy 3458 AGCTGAAGCGGAGAGCTTCTGCAAGCTCAATGATCTGAGAGAGCAAGCTATG 3517
Db 2215 AGCTGAAGCGGAGAGCTTCTGCAAGCTCAATGATCTGAGAGAGCAAGCTATG 2274
Qy 3518 TTGAATGATGCGCGGAGCTTACAGAGAGCTGAGAGCTGAAGAGCTCAAGAG 3577
Db 2275 TTGAATGATGCGCGGAGCTTACAGAGAGCTGAGAGCTGAAGAGCTCAAGAG 2334
Qy 2275 TGAAGATGAGCGCGGAGCTTACAGAGAGCTGAGAGCTGAAGAGCTCAAGAG 2394
Db 2335 TGAAGATGAGCGCGGAGCTTACAGAGAGCTGAGAGCTGAAGAGCTCAAGAG 2394
Qy 3638 TCCGCTGATCTGAGAGCTGCAAGAGCTTGAAGCTGAGAGCTGATTAAGAGAGAA 3697
Db 2395 TCCGCTGATCTGAGAGCTGCAAGAGCTTGAAGCTGAGAGCTGATTAAGAGAGAA 2454
Qy 3698 GAAATGATGAGAGTCAAGCTGAGAGAGCTTCAAGCTTCAATCTGATGAGAGAG 3757
Db 2455 GAAATGATGAGAGTCAAGCTGAGAGAGCTTCAAGCTTCAATCTGATGAGAGAG 2514
Qy 3758 AATGAGAGAGCTATTTTCTGAACCAAGCTGATTTTCTGAAGCCAAATG 3817
Db 2515 AATGAGAGAGCTATTTTCTGAACCAAGCTGATTTTCTGAAGCCAAATG 2574
Qy 3818 ACCAAGCTGATGAGAGAGAGCTTCTGAGAGTCAAGCTGAGAGTGGCCCTG 3877
Db 2575 ACCAAGCTGATGAGAGAGAGCTTCTGAGAGTCAAGCTGAGAGTGGCCCTG 2634
Qy 3878 AGAGAGAGAGAGCTGCTGAGAGAGCTGAGAGAGCTTCAAGAGAGCTGAGAG 3937
Db 2635 AGAGAGAGAGAGCTGCTGAGAGAGCTGAGAGAGCTTCAAGAGAGCTGAGAG 2694
Qy 3938 TCCGCTGCGCGGAGAGAGCTGCGCCACCGCAAGAGCAAGCAAGCAAGCAAG 3997
Db 2695 TCCGCTGCGCGGAGAGAGCTGCGCCACCGCAAGAGCAAGCAAGCAAGCAAG 2754
Qy 3998 CGCCAGAGAGAGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 4057
Db 2755 CGCCAGAGAGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 2814
Qy 4058 AGCCAGAGAGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 4117
Db 2815 AGCCAGAGAGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 2874
Qy 4118 CAGAGAGATTTAGTGGAGCTTGAAGAGAGAGTCCAGAGTCCAGAGTCCAGAG 4177
Db 2875 CAGAGAGATTTAGTGGAGCTTGAAGAGAGAGTCCAGAGTCCAGAGTCCAGAG 2934
Qy 4178 AGTGAAGAGTGAAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 4237
Db 2935 AGTGAAGAGTGAAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 2994
Qy 4238 GAGCGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 4297
Db 2995 GAGCGAGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAG 3054

QY	4298	CGTTGACAGCCACCCTGGCGCTTGCTGCCTGAAATATGACCAACACTTACCGAGGCTTCT	4353
Db	3055	CGTTGCCCGCTACCTTGTTGGCCCTGCCAGCTGAATATGCCACACACTTCACTGAGGCTTCT	3114
QY	4358	GCCGTGACAAATATGAATCTCCCAAGGCTTCCAGACCCAGAGAGGCCGACGACGCTTGCACC	4417
Db	3115	GCCGGGACAAATATGAATCTCCCGGGGCTTCCAGAGCAAGAGGCTTGGCAGACGCTTGCACC	3174
QY	4418	TGGAAAGGTGATGAAAGTGGCCACAGAAATTAACAACGAGGACAGAAAGGCTTGGGACAG	4477
Db	3175	TGGAAAGGTGATGAAAGTGCCCAAAATTAACAAACGAGGACAGAAAGGCTTGGGACAG	3234
QY	4478	AGTACATTGCTCTGAGAGGATCAAAAGTCTCATTTATGACATATGACCAGAGAGCTG	4537
Db	3235	AGTACATTGCTCTGAGAGGGGTCAAAAGTCTCATTTATGACATATGACCAGAGAGCTG	3299
QY	4538	GACAGAGCCCGGTGGAAGATTTGAGCTGTGCTTTCCGACGGGGATGATCTATTCATG	4597
Db	3295	GACAGAGCCCGGTGGAAGATTTGAGCTGTGCTTTCCGACGGGGATGATCTATTCATG	3355
QY	4598	GTGCGCGTGTGTGCTTCCGAACTGGCAATPACAGCAAGACAGATGTCCCATACATACGA	4655
Db	3355	GTGCGCGTGTGTGCTTCAAGACTTGCATAATACCGCCAAAGCAGATGTCCCATACCTTGA	3414
QY	4658	AGATGAATCTCACCCCGACACCAACTGTGCTGGCCCGGAGAAACCTCTACTTGCTAGCTC	4717
Db	3415	AGATGAATCTCACCCACACACCAACTGTGCTGGCCGGAGAAACCTCTACTTGCTAGCAC	3474
QY	4718	CCAGCTTCCCTGACAAACAGCGCTGGGTCAACGCTTGAATCAATGTGCGAGGTGGGA	4777
Db	3475	CCAGCTTCCCGACAGACAGCGCTGGGTCAACGCTTGAATCTGTGCTGCGAGGTGGGA	3533
QY	4778	GAGTTTCTAGGAGAAAAACAGAGGTGATGTCTAACTGCTTGGAAACTCCCTGCTGAAC	4837
Db	3535	GAGTTTCTAGGAGAAAAAGGCCAGACCGATGTCTAATTACTTGGAACCTCTGCTGAAC	3599
QY	4838	TGGAAGGTGATGACCGCTGTAGACATGAACTGACGCTGCCCTTCACTAGTACCAAGGTGTGT	4897
Db	3595	TGGAAGGTGATGACCGGCTTGAACAATGAACTGACGCTGCCCTTCACTAGTACCAAGGTGTGC	3655
QY	4898	TGCTGGGACACGAGAGAGGCTCTTGCGCCCTGAATGTCTTGAATAACTCCCTAACCCATG	4957
Db	3655	TGCTGGGACACGAGAGAGAGGCTCTTACGCGCTTGAATGTCTTGAATAACTCCTTAACCCACA	3714
QY	4958	TCCCAAGAATTGAGACAGCTTCCAAATTTATATTAACAAGACTGGAAGGCTATCTCA	5017
Db	3715	TCCCAAGGATTGGCCAGACTTCTCCAAATTTATCATATCAAGGACTGGAAGGCTGCTCA	3774
QY	5018	TGATAGCAGAGAAAGACGGGACACTGTGTCTTGTGAAGTGAAGAAAGTGAACAGTCC	5077
Db	3775	TGATAGCAGGGGAAGAGCGGGCTCTGTGCTGTGTGAAGTGAAGAGTGAACAGTCCC	3833
QY	5078	TGGCCCACTCCACCTGCTGCTGCCAGCCGACATCTCAACCAACTTTTGAAGCTGTCA	5137
Db	3835	TGGCCCACTCACACTGCTGCTGCCAGCCGACATCTCCCCCAAAATTCGAAGCGCTCA	3894
QY	5138	AGGGGTGACACTGTTTGGGGAGGCAAGATTGAAGAGGGTCTGCATCTGTGAAGCA	5197
Db	3895	AAGGTGCACTGTTGCTGCTGTGGCAAGATGGAACAAGGCTTGCAATCTGCGCGCTGA	3954
QY	5198	TGCCCAGCAAAAGTGTCTCAATTCCTCCGCTCAACGAAACCTCAGCAAAATACGATCCGGA	5257
Db	3955	TGCCCAGCAAAAGTGTCTCAATTCCTCCGCTCAATGACAACTCAGCAAAATACGCAACGGA	4014
QY	5258	AAGAGATGGAACCTTCAGAGCCCTGACAGTATCACTTACCAATTAACAGTATCTCA	5317
Db	4015	AAGAGATGGAACCTTCAGAGCCCTGACAGTATCACTTACCAACATACAGCAATCTCA	4074
QY	5318	TTGGAACCAATTAATTTCAACGAATTCGACATGAACAGTATCAACGCTTCAGAGAAATTCCTGG	5377
Db	4075	TTGGAACCAATTAATTTCAATGATGACATGAACAGTATCAACGCTTGAATGATGTTCTCGG	4134
QY	5378	ATTAGAAATGACCAATTCCTTGGACCTGCTGTGTTTGGCGCTCTTCCAAACAGCTTCCCTG	5437

Db	4135	ACAGGAACGACCAATTCCTTGGCACCTGCTGTGGCTCCTCGTCCACAGCTTCCCTG	4194
OY	5338	TCTCAATCGTCAGGTGAACAGCGGACAGGACGAGAGATCTTGTGTTTCCACG	5497
Db	4195	TCTCATTTGTCAGGCGCAACAGCGCCCGGACAGCAAGAAATATCTGCTGTCTTCACG	4254
OY	5498	AATTGGAGGTTCGTGGATTTCTTACGGAGAGAGTACCGCACAGACGATCTCAAGTGA	5557
Db	4255	AATTGGAGGTTCGTGGATTTCTTACGGAGAGAGTACCGCACAGATGATCTCAAGTGA	4314
OY	5588	GTGCTTACCTTTGGCTTGGCTTGCCTACAGAGAACCTTATCTGTTGTACACCACTTCACT	5617
Db	4315	GTGCTTACCTTTGGCTTGGCTTGGCTTGCCTACAGAGAACCTTATCTGTTGTACCTTCACT	4374
OY	5618	CATCTGAAGTAAATTAAGATCCAGGCACGCTCTTACAGAGAGAACCCCTGCGGACGTAAC	5677
Db	4375	CCCTGGAAGTCAATTAAGATCCAGGCACGATCTCTACTGCGGAGGCTGCGCGACATATC	4434
OY	5678	TGACATTCGCGAATCCGCGGCTACCTGGAGCCCTGECATTTCTCTAGAGAGCATTAATCTAG	5737
Db	4435	TGGAATTTCCAAACCTTCGCTACTGAGGCCCTGCGATTTCTCTCGGAGCATTAATCTAG	4494
OY	5738	CGTCTCTAATACAGAGATTAATTAAGGATCAATTTGCTGCAAGGAACTCTGTGAAGAGT	5797
Db	4495	CCTCTCTAATACAGAGATTAATTAAGGATCAATTTGCTGCAAGGAACTCTGTGAAGAGT	4554
OY	5798	CCGCGCATCGAACACACCGGGGCCCCGTCACCTCCGACAGAGCCCAACAGAGAGAGCC	5857
Db	4555	CAGGACCTGAGACACACCGGTCCTTCACCTTCGAGAGAGCCCAACAGAGAGAGCC	4614
OY	5858	CACCCACGTACCAAGACACATCAACAAAGGCGTGAGCTCCAGCCACAGCGCCCGCAAG	5917
Db	4615	CACCAACATACCAAGACACATCAACAAAGGCGTGAGCTCCAGCCACAGCGCCCGCAAG	4674
OY	5918	GCCCCAGCCACCCCGGAGAGCCCAAGCACACCCACCGCTAAC-----GCGAGGGCGCGA	5971
Db	4675	GCCCCAGCCACCCCGGAGAGCCCAAGCACACCCACCGCTAAC-----GCGAGGGCGCGA	4734
OY	5972	CCGAGCTGCGCAGAGGACAAAGTCTCTGAGCGGCCCCCTCGAGAGCGAGAAAGTCCCCGAGC	6031
Db	4735	CAGAGCTGCGCAGAGGACAAAGTCTCTGAGCGGCCCCCTCGAGAGCGGAGAAAGTCCCCGAGC	4794
OY	6032	GGATTACTGACACCGGAGAGAGCGGTCCCCCGGAGGCTGTTGAAGACACGACGAGG	6091
Db	4795	GAAAGCTCAGACACTAGAGAGAGAGGCGTCCCCGAGGAGACTGTTGAAGACACGACGAGG	4854
OY	6092	GCGGAGTGCCTCGCGAGAGCGGTGAGACCCCGGCTGTCCGAGTGAACAAAGTCTGGGAGC	6151
Db	4855	GCGGAGTGCCTCGCGAGAGCGGTGAGACCCCGGCTGTCCGAGTGAACAAAGTCTGGGAGC	4914
OY	6152	AGTCTTCAGTATAA 6165	
Db	4915	AGTCTTCAGTATAA 4928	
RESULT 15			
LOCUS	MMUC3904		
DEFINITION	Mus musculus c17ron, putative rho/rac effector, mRNA, complete cds.		
ACCESSION	U3904		
VERSION	U3904.1		
KEYWORDS	signal transduction; rho; coiled-coil; leucine-zipper; plectstrin homology; zinc finger.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Medanle,P., Furuyashiki,T., Reid,T., Ishizaki,T., Watanabe,G., Morita,N. and Narumiyu,S.		
TITLE	A novel partner for the GTP-bound forms of rho and rac		
JOURNAL	FEBS Lett. 377 (2), 243-248 (1995)		

MEDLINE	96128238
PUBMED	8543060
REFERENCE	2 (bases 1 to 5019)

AUTHORS
 Madaule, P., Fukuyasu, K.I., Reig, J., Ishizaki, I., Matsumae, G.,
 Moril, N. and Narmiyu, S.
 TITLE
 Direct Subunit
 Submitted (02-NOV-1995) Pascal Madaule, Department of Pharmacology
 Faculty of Medicine, Kyoto University, Yoshida, Konoe, Sakyo-ku,
 Kyoto, 606, Japan
 FEATURES
 source
 1. 5019
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="brain"
 /dev_stage="adult"
 1. 204
 /note="Multiple alternative splices affect the N-terminal
 region. The 5' end may be replaced by another sequence,
 resulting in a protein with an N-terminal extension"
 182..4975
 /note="binds to the GTP-bound forms of p21 rho and rac,
 according to two hybrid system and overlay assay; putative
 rho/rac effector"
 /codon_start=1
 /product="citrion"
 /protein_id="AAC52344.1"
 /db_xref="GI:10797344"
 /translation="MLDEBAMRMEQENTRLHRRVSEYVALISQKVELKASTQSL
 EDLALVITECSLSKSLSEQAARMEVSQDEDAQLADHIDREGRKQVETKEEYQAQ
 EMRLMMNOLEDIVARSRSDDLYSELSRESRLAAEFPRKMAQCHIKMKAKDQK
 EYGEVSEGLIKVAEQOLQIOELQEKERAVASVETALQNLIRQAKAEALDELKLI
 NNDESGEIKKKLVAEERRRSLEIKYKRLTMERRNRRLKQDIOKSHQIOQMDKY
 LLEEHRAEYSAQHLVHLKQKQHYEKKYTLQDQIKQIDAPESILMMGSHPE
 BAEKKKILSEQKAMITNMDSKISLEGRIVELSEANKLAANSLETTQNMQAQEM
 SELRQKQVLTLETQAGLEAONRKLSEBQLEKISHODSDKSLANSLETLRLVSLHEH
 QKLELRQLTLETQSLQESRQTLQAARAALESQKXETLETTLEAEERIOA
 THARPEIQKPEPALRNSCTVITDEEOLNQTENMALNNQNFYLSKQDASGAND
 IOVRESVDHLRERTITEREMOTQOKQMEALKTTCTMLEQVLDLALNDLLEKEH
 ONMANSYVLDGEKQFEQREVRLEQRLDTEKQSRARADQITSRQVVALYEHKKA
 ILAQALKEQKLEKESLSDKLDLBEKKHMLBNMARSLOQKLETRRLKQNLLEQ
 KQAQNDQDKKHTFRLTQLOELADRLDLKTERSDLEIQLNTQVLYSEHVRNNEG
 ISQQTLDLDFLOAKKDQPAKKKVPQLYNEILKALEKERAQCALEBALQKRIELR
 ABEBAHRKATDHPSPTDPATAROOIMASIVSPHROFASMLAPSSRKESSCT
 AEESSRLKERMHNIPIHRFVNLMMARATKACVLDYTFPGSGLCEQWCKESST
 TGLPATCGPAEYATHPEAFGRDKNMSPGLOSPSSSLHEGMMKVRNNKRRQDQ
 MWRKYVLEGSKULIYDNEAREAGORPGEELCPDQVSTIGAVGASLEIANTAKA
 VPTILMESHPPTTCMPGRTIYILASPDPKQRYTALSEVAGVSEKREKALPAK
 LGSULKLTMDRDMQCTLPSPDVAVVGEBSLAVLANTLNSHTIGAVQITL
 ILDKLEKLTLAGGERALCIVDVKVQSLAQLADPQVNSNIPIEAGQCHLPAQA
 KIENSICLCAAMPKSVKILIRYNDMLSKCIKRELTESEPSCICHFNPNYILITGKVF
 EIDMKQYLTLELDKNDGSLAPAVFASNSNFPVSIYOVASAGOREYILCPHEGVG
 VDSYGRSRTDLDKWSRLPLAFAEPIFYTHNNSLEVEIAGAESLSGPARATLE
 PDNRYTALISGATYILASQVODKLRVYCCKGNLYVSGTEGRQVSTSRSSPNKRG
 PYNTEHTIRVASSPAPBPBGSHPRBPTPRTRIDREGRTELRRDSPPRLERKSI
 GRMLSTRRSRSGRLFEEDSSKRGRLPAQAVPTPLSYNKKWDSSV"
 887..1012

BASE COUNT	1399 a	1350 c	1466 g	804 t	/note="this exon is dispensable"
ORIGIN					
Query Match		65.2%	Score 4018.2;	DB 10;	Length 5019;
Best Local Similarity		90.3%;	Pred. No. 0;		
Matches 4309;		Conservative 0;	Mismatches 458;	Indels 6;	Gaps 1
QY	1399	AAAGATGAGCAGGAAATGACCCCGTTTACATGGAGAGTGTTCAGAGGTGGAGGCTGTGCTT	1456		
Db	203	ATATGAGCAGAGGAATGACCCCGTTTACATGCGCAGAGTGCAGAGGTGGAGGCTGTGCTT	262		
QY	1459	AGTCAGAGGAGGTGAGCTGAGAGGCCCTGAGACTGAGAGATCCCTCCGTGAGCAGAGAC	1511		
Db	263	AGTCAGAGGAGGTGAGCTGAGAGGCCCTTCGTGACTCAGAGATCCCTCCGTGAGCAGAGAC	322		

QY	1519	TTTCTCAACCTACATCAACAGATGAGTAGCTTTAAAGGAAGTTTGGACCAAGCAGGATG	1578
Db	323	CTTCTCACTCTACATCAACAGATGAGTAGCTTTAAAGGAAGTTTGGACCAAGCAGGATG	382
QY	1579	GAGGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCATGATATCAGAGACAGAGC	1638
Db	383	GAGGTGTCCAGAGAGATGACAAAGCTCTGAGCTTCTCACAGCATCCAGAGACAGAGC	442
QY	1639	CGAAGAGCTCCAGAAATATCAAAAGACAGAGATACCAAGGCTCAAGTGGAAAGAAATGAGGTTG	1698
Db	443	CGAAGAGCTCCAGAGAGATATCAAGAGACAGAGATACCAAGGCTCAAGTGGAGAGATGAGGCTG	502
QY	1699	ATGATGAATCAGTTGTGAAGAGAGATCTTGTCTCAGACAAAGAAAGACGAGTGCATCTCTAGAA	1758
Db	503	ATGATGAATCAGCTGTGAAGAGAGATCTGTGTCTCAGCCGACAGACGACGATCTCTACAGAG	562
QY	1759	TCTGAGCTGAGAGAGTCTCGGCTTGTCTGTGAAGAAATTCAGACGAGAAACGACAGAAATGT	1818
Db	563	TCTGAGCTGAGAGAGTCTCGGCTTGTCTGTGAAGAAATTCAGACGAGAAACGAGAAATGT	622
QY	1819	CAGATATTAACGTGTGAAGAGCTTAAGAGATCAAGAGGAGCTTAAGTGGAGAGAAATATGCGAAA	1878
Db	623	CAGACAAATCTGATGAAGGCTTAAGAGACCAAGAGGAGCTTAAGTGGAGAGAAATATTCGAAA	682
QY	1879	CTGAGAGAGATCAATGCTGTAGACAGACAGCTCAAAATTCAGAGAGCTCCAAGAGAACTGGAG	1938
Db	683	CTGAGAGAGATCAATGCTGTAGACAGACAGCTCAAGATTCAGAGAGCTCCAAGAGAACTGGAG	742
QY	1939	AAGGCTTTAAAGGCGACACAGAGGCGACCGAGGTGTGACAGATATATCCGACAGGCAAG	1998
Db	743	AAGGCTTTAAAGGCGACACAGAGGCGACCGAGGTCTTGACAGAAATATCCGACAGGCAAG	802
QY	1999	GAGGAGCGCCGAGAGAGGAGCTGTGAGAGACCTGCAGAACCGAGAGAGATTTCTTCTGAAAGGATC	2058
Db	803	GAGGAGCGAGGCGGAGAGCTAGAGAAAGCTACACACCGAGAAAGATCTCTCGAAGGATC	862
QY	2059	AGAAAGAAAGCTGTGTGAAGCTGTAGAGAAACCGCCCATTTCTGTGAGAAACAGCTAAAGAGA	2118
Db	863	AAAAAGAAAGCTGTGTGAAGGCGAGAGAAACCGCCCATCTCTGTGAGAAACAGCTAAAGAGA	922
QY	2119	CTAAGAGCAATGTAGAGCGTTAGAGAAACAGACTGAGAGATGATCATTCAGACAAAAATCCCAA	2178
Db	923	CTAAGAGCAATGTAGAGCGTTAGAGAGAACAGATGAGAGATGATCATTCAGACAAAGATCCGAA	982
QY	2179	CAGATCCAGCAGATGTGCTATTAATTTCTGAGAGCTCGAAGAGAAACATTCGGAGAGCCCAA	2238
Db	983	CAGATCCAGCAGATGTGCTATTAATTTCTGAGAGCTCGAAGAGAGAAACATTCGGAGAGCTCAG	1042
QY	2239	GTCTCAGCCCGACACTAGAGGTGCACCTGAAACAGAAACAGACGACATATGAGAGAAAG	2298
Db	1043	GTCTCAGCTCAACACTTGAAGATGACACTTGAACAGAGAGAACGCACTACAGAGAGAAAG	1102
QY	2299	ATTAAAGGTGTGGAACATTCAGATCAATTAAGAAAGACCTGTGCTGACAGAGAGACACTGSAAGC	2358
Db	1103	ATCAAGGATTTGGACATTCAGATTAAGAAAGAGCCTGTGCCAGACAGAGAGAGCTGTGAGAAC	1162
QY	2359	ATGATTCAGAGACACAGAGAGAGAGGCGCCATGAGAGAGGCAAAATTTCTACGCAACAGAG	2418
Db	1163	ATGATTCAGAGACACAGAGAGAGAGGCGCCATGAGAGAGGCAAAATTTCTACGCAACAGAG	1222
QY	2419	GCGATGATCAATGTGTAATGAGATTCAGAAATCAGATCCCTGGAGACAGAGATTTGTGGAATG	2478
Db	1223	GCGATGATCAACGAGATGAGATTCGAAATCCGATCCCTGGAGACAGAGATTTGTGAGCTG	1282
QY	2479	TCTGAGCGCAATTAACCTTGACAGCAATAGCAGTCTTTTATCCCAAGAGAAATATGAGGCTC	2538
Db	1283	TCTGAGCGCAACAGAGCTTGGCGGCAAAACAGCAGTCTCTTCAACCCAGAGAGAAATATGAGGCTC	1342
QY	2539	CAAGAGAGATGATTTCTGGAATCAGGCAACAGAAATTTTAACTGTGAGACACAGGCTGGG	2598
Db	1343	CAAGAGAGATGATCTCAGAACTCAGGCAAGAAATTTTAACTGTGAGACACGAGCTGGG	1402
QY	2599	AAGTTTGAAGGCCAGAACGAGAACTTGAGAGACAGCTTGAGAGAAATCAGCCACAGAGC	2658

Db 1403 AAGCTGAGGCCCCAGAACCGAAAGCTGGAGAGCAACTGGAGAAAGATCAAGCCACCAAGAT 1462
Qy 2659 CACAGTGAACAGAAATCGCTGCTGGAATTGGAGACAAGATTGCGGAGGCTCACTTATGAG 2718
Db 1463 CACAGTGAACAGAGTGGCTGCTGAGCTGGAAACAAGGCTGAGGAGGTCAGCCCTGGAG 1522
Qy 2719 CACGAGAGCAGAAATTGAGGCTCAGGCCCAGCTCAAGAGCTTACAGCTTCTCCCTGAG 2778
Db 1523 CACGAGAGCAGAAAGCTGAGGCTGAAAGCGGAGCTTACGAGGCTGAGCTGCTCCCTGAG 1582
Qy 2779 GAGCGGAGTCAACAGTTGACAGCCCTGACAGGCTGACGGGCGGCTTGGAGAGCCGCTT 2838
Db 1583 GAGCGGAGTCCACAGCTGACGCGCCCTGACAGCCGCGCGGAGGCTTGGAGAGCCAGCTTC 1642
Qy 2839 CCGCAGGCGAAGACAGAGCTGGAAGACACACAGAAAGCTGAAGAGAGATCCAGGCA 2898
Db 1643 CCGCAGGCGAAGACAGAGCTGGAAGAGACACCGCGGAAGCGGAGAGAGATCCAGGCG 1702
Qy 2899 CTACGCGGACATAGAGATGAATCCAGCGGCAATTGATGCTCTTCTGTAACAGCTTACT 2958
Db 1703 CTACGCGGACATCGAGATGAATCCAGCGCAATTGATGCTCTTCTGTAACAGCTTACT 1762
Qy 2959 GTATCAACAGACCTGAGAGGAGGCTAAACAGCTGACCGAGGACACACGCTGAACCTCAAC 3018
Db 1763 GTATCAACAGACCTGAGAGGAGGCTGAACAGCTGACCGAGGACACACGCTGACAC 1822
Qy 3019 AACCAAACTTTACTTCTGCTCAACACATCTGATGAGGCTTCTGCGCCCAACGACGAT 3078
Db 1823 AACCAAACTTTACTTCTGCTCAACACATCTGATGAGGCTTCTGCGCGCCCAATGACGAT 1882
Qy 3079 GTACAACTGCGAAGTGAAGTGAACCATCTCGCGCGGAGATGACGGAACGAGATGACG 3138
Db 1883 GTGACGCTGCGAAGTGAAGTGAACCATCTCGCGCGGAGATGACGGAAGGAGATGACG 1942
Qy 3139 CTTACCAAGCAGAACCAACGATGAGGCTCTGAAGACACAGCTGACCAATGCTGGAGAA 3198
Db 1943 CTTACCAAGCAGAACCAACGATGAGGCTCTGAAGACACAGCTGACCAATGCTGGAGAA 2002
Qy 3199 CAGGCTCATGATTTGGAGGCTCTTAACGATGAGCTTGAAGAAAGAGCGGACGAGTGGAG 3258
Db 2003 CAGGCTCATGATTTGGAGGCTCTTAACGATGAGCTTGAAGAAAGAGCGGACGAGTGGAG 2062
Qy 3259 GCGTGAAGAGGCTCTGAGGATGAGAAATCCGATTGAGTGTGCGGCTTGAAGAGCTG 3318
Db 2063 GCGTGAAGAGGCTCTGAGGATGAGAAATCCGATTGAGTGTGCGGCTTGAAGAGCTG 2122
Qy 3319 CAGAGGATGCTGACACCGAGAAACAGAGCGGCGAGAGCGATCAGCGATCCAGAG 3378
Db 2123 CAGAGGATGCTGACACCGAGAAACAGAGCGGCGTAAAGGCGATCAGCGATCCAGAG 2182
Qy 3379 TCTGCGCAGGTGAGTGGAGCTGGCAGTGAAGAGCACAAGGCTGAGATTTCTGCTTGGAG 3438
Db 2183 TCTGCGCAGGTGAGTGGAGCTGGCAGTGAAGAGCACAAGGCTGAGATTTCTGCTTGGAG 2242
Qy 3439 CAGGCTCTCAAGAGCAGAAAGCTGAAGGCGGAGGCTCTGACACAGCTCAATGACCTG 3498
Db 2243 CAGGCTCTCAAGAGCAGAAAGCTGAAGGCGGAGGCTCTGACACAGCTCAATGACCTG 2302
Qy 3499 GAGAAAGAGCATGCTTATGCTTGAATGAAATCCGAAAGCTTACAGAGAAAGCTGAGACT 3558
Db 2303 GAGAAAGAGCATGCTTATGCTTGAATGAAATCCGAGGCTTACAGAGAAAGCTGAGACA 2362
Qy 3559 GAAAGAGAGCTCAACAGAGGCTTCTGAAAGAGCAGCAATTTACAGCAGCAGATGGAC 3618
Db 2363 GAAAGAGAGCTCAACAGAGGCTTCTGAAAGAGCAGCAATTTACAGCAGCAGATGGAC 2422
Qy 3619 CTGCAAGAAATCAATTTTCCGTCTGACTCAAGGACTGCAAGAGGCTTATGATCGGCT 3678
Db 2423 CTGCAAGAAATCAATTTTCCGTCTGACTCAAGGACTGCAAGAGGCTTATGATCGGCT 2482
Qy 3679 GATTACTGAAGAGAAAGATGACTTGGAGTATGAGCTGGAAAATTTCAAGTTCTC 3738

Db 2483 GATTCTGTAAGACAGAAAGAGCGACTGAGTACCACTGGAAAAATTCAGGTTCTC 2542
Qy 3739 TATTCTCATGAAAGGTGAAAAATGAAAGGCACTATTCTCAACAAACCAATCATTTGAT 3798
Db 2543 TACTCTCAGAGAAAGTAAAAATGAAAGGCAATCTTCAAGAAACCAAACTCATTTGAT 2602
Qy 3799 TTTCTGCAAGCCAAATGAGCCACCTGCTAAAAAGAAAAAGTTCTCTGCAATCAAT 3858
Db 2603 TTTCTGCAAGCCAAATGAGCCACCTGCTAAAAAGAAAAAGTTCTCTGCAATCAAT 2662
Qy 3859 GAGCTGAAGCTGGCTTGGAGAAAGAAAGCTCGCTGTCAGAGCTTAAGAGAGCCCTT 3918
Db 2663 GAGCTGAAGCTGGCTTGGAGAAAGAAAGCTCGCTGTCAGAGCTTAAGAGAGCCCTT 2722
Qy 3919 CAGAAAGACCGGATGAGGCTCGGCTCCGCGGAGGAGAGCTGCCCCACCGCAAGCAACG 3978
Db 2723 CAGAAAGACCGGATGAGGCTCGGCTCCGCGGAGGAGAGCTGCCCCACCGCAAGCAACG 2782
Qy 3979 GACCAACCAACCAATCCACGCGCAGCAGCCGAGGAGCAGATCGCATGTCGCGCATC 4038
Db 2783 GACCAACCGCAGCCGCTTACGCGCAGCAGCTGAGGAGCAGATCGCATGTCGCGCAT 2842
Qy 4039 GTGCGGTGGCCAGAGCAGCCAGCTGCTGATGAGCTGCTGAGCCCGCATTCAGCCGC 4098
Db 2843 GTGCGGTGGCCAGAGCAGCCAGCTGCTGATGAGCTGCTGAGCCCGCATTCAGCCGC 2902
Qy 4099 AGAAAGAGTCTTCAACTCCAGAGAAATTTGTGGCGGCTTAAAGAAAGCATGACACAC 4158
Db 2903 AGAAAGAGTCTTCAACTCCAGAGAAATTTGTGGCGGCTTAAAGAAAGCATGACACAC 2962
Qy 4159 AATATCTCTCAACGATTAACGATGAGCTGAACATGCAAGCCACAAAGTGTGCTGTGT 4218
Db 2963 AATATCTCTCAACGATTAACGATGAGCTGAACATGCAAGCCACAAAGTGTGCTGTGT 3022
Qy 4219 CTGATTAACGTCACATTTTGAAGCCAGGATTCAAATGCTTCTGAAATGTGAGTGTGT 4278
Db 3023 CTGATTAACGTCACATTTTGAAGCCAGGATTCAAATGCTTCTGAAATGTGAGTGTGT 3082
Qy 4279 CACCCCAAGTCTCCACGCTGTCGACGACCTGCGGCTTGGCTGGAATATGACCA 4338
Db 3083 CACCCCAAGTCTCCACGCTGTCGACGACCTGCGGCTTGGCTGGAATATGACCA 3142
Qy 4339 CACTTCAACGAGGCTTCTGCGCTGACCAAAATGAATCTCCGAGCTCTCAGACCAAGAG 4398
Db 3143 CACTTCAACGAGGCTTCTGCGCGGACAAATGAATCTCCGAGGCTCTCAGACCAAGAG 3202
Qy 4339 CCGAGCAGACCTTGCACCTGGAAGGCTGATGAAGGTCGCCAGAAATPAACAGAGGA 4458
Db 3203 CCGAGCAGACCTTGCACCTGGAAGGCTGATGAAGGTCGCCAGAAATPAACAGAGGA 3262
Qy 4459 CAGCAAGGCTGGACAGGAAGTACATGTCCTGAGAGGATCAAAAGTCTCATTTATGAC 4518
Db 3263 CAGCAAGGCTGGACAGGAAGTACATGTCCTGAGAGGATCAAAAGTCTCATTTATGAC 3322
Qy 4519 AATGAAGCAGAGAGCTGACAGAGGCGGTGGAAGAAATTTGAGCTGTGCTTCCGAC 4578
Db 3323 AATGAAGCAGAGAGCTGACAGAGGCGGTGGAAGAAATTTGAGCTGTGCTTCCGAC 3382
Qy 4579 GGGAGATGATTTATTTATGTCGCTGCTGCTTCCGAACTGCGAAATPAACGCCAAGCA 4638
Db 3383 GGGAGATGATTTATTTATGTCGCTGCTGCTTCCGAACTGCGAAATPAACGCCAAGCA 3442
Qy 4639 GATGTCCATACATACATGAGATGGAATTCACCCGACACCACTGTCGCGCGGAGGA 4698
Db 3443 GATGTCCATACATACATGAGATGGAATTCACCCGACACCACTGTCGCGCGGAGGA 3502
Qy 4699 ACCCTTCACTTGTAGCTCCAGCTTCCGTAACAAACAGCGCTGGGTCAACGCTTGA 4758
Db 3503 ACCCTTCACTTGTAGCTCCAGCTTCCGTAACAAACAGCGCTGGGTCAACGCTTGA 3562
Qy 4759 TCAAGTTGCGAGGTGGAAGTTCAGGGGAAAAACCAAGAGTATGCTAACTGCTT 4818
Db 3563 TCTGTCTCCAGGTGGAGAGTTTCAAGGAAAAAGGCCAGAACCGATGCTTAATTA 3622

QY 4819 GGAAGTCCCTGTAAGTGAAGTATGACCGTGTAGACATGATGACGCTGCCC 4878
DB 3623 GGAAGTCTCTGCTGGAAGTGAAGGCGATGACCGGTTGACATGATGATGACCTGCCC 3682
QY 4879 TTGAGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4938
DB 3683 TTGAGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3742
QY 4939 AAAAGTCCCTGTAAGTGAAGTATGACCGTGTAGACATGATGACGCTGCCC 4998
DB 3743 AAAAGTCTCTGTAAGTGAAGTATGACCGGTTGACATGATGATGACCTGCCC 3802
QY 4999 GACCTGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5058
DB 3803 GACCTGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3862
QY 5059 AAGGAAGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5118
DB 3863 AAGGAAGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3922
QY 5119 AACATTTTGAAGTCTGTAAGTGAAGTATGACCGTGTAGACATGATGACGCTGCCC 5178
DB 3923 AACATTTTGAAGTCTGTAAGTGAAGTATGACCGGTTGACATGATGATGACCTGCCC 3982
QY 5179 CTGTCATCTGTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5238
DB 3983 CTGTCATCTGTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4042
QY 5239 AGCAATATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5298
DB 4043 AGCAATATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4102
QY 5299 ACCAATTAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5358
DB 4103 ACCAATTAAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4162
QY 5359 ACGCTGAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5418
DB 4163 ACGCTGAGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4222
QY 5419 TCTTCAAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5478
DB 4223 TCTTCAAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4282
QY 5479 TACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5538
DB 4283 TACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4342
QY 5539 ACAGAGATCTGTAAGTGAAGTATGACCGTGTAGACATGATGACGCTGCCC 5598
DB 4343 ACAGAGATCTGTAAGTGAAGTATGACCGGTTGACATGATGATGACCTGCCC 4402
QY 5599 TTTGTGACCACTTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5658
DB 4403 TTTGTGACCACTTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4462
QY 5659 ACCCTGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5718
DB 4463 ACCCTGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4522
QY 5719 TCAAGAGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5778
DB 4523 TCAAGAGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4582
QY 5779 GGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5838
DB 4583 GGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4642
QY 5839 AGCCCAAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5898
DB 4643 AGCCCAAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4702

QY 5899 AGCCCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5958
DB 4703 AGCCCAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4762
QY 5959 C-----CGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6012
DB 4763 CGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4822
QY 6013 CGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6072
DB 4823 CGAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4882
QY 6073 TTTGAAGCAGAGCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6132
DB 4883 TTTGAAGCAGAGCAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4942
QY 6133 GTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6165
DB 4943 GTTAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4975

Search completed: November 15, 2003, 12:14:23
Job time : 14501.3 secs